Webinar Series

Data Use Skills
Featuring Data from Natural History Collections

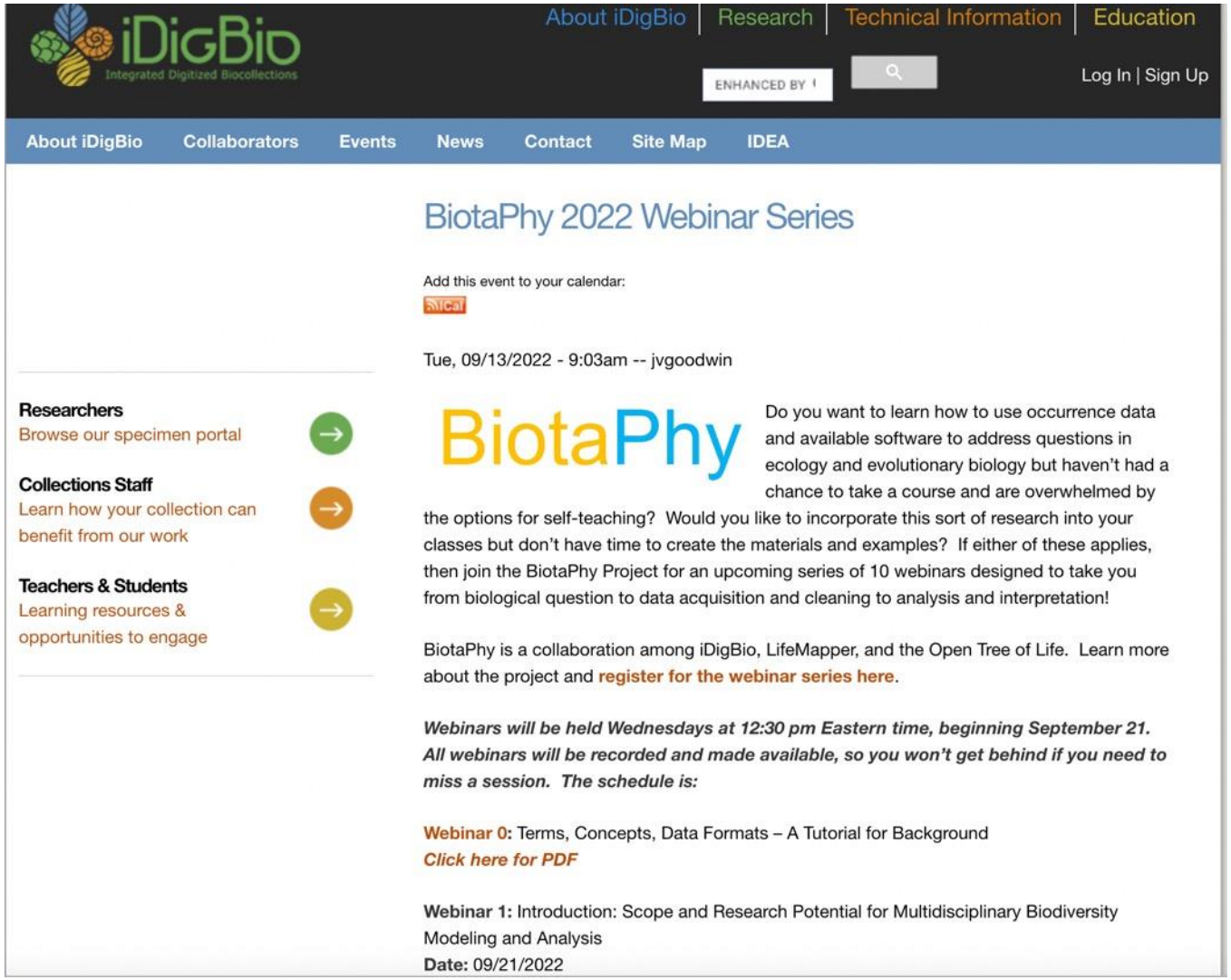
September 21-November 30, 2022

https://www.idigbio.org/content/biotaphy-2022-webinar-series

iDigBio:

https://www.idigbio.org/content/biotaphy-2022-webinar-series

iDigBio.org



Thank You

Maria Cortez
Aimee Stewart

Jill Goodwin
Gil Nelson

Webinar 6

Species Distribution Models: What are they? How to Create One?

Goals

Create Species Distribution Models using Maxent using occurrence data with the minimum number of points defined in the configuration file or the Rare Species Model algorithm for data without the required minimum number of points.

Learning Objectives

Biological Objectives:

- **Species Distribution Models**
- **Fundamental and realized niche**
- **Applications**

Technical Objectives :

- **Input occurrence records, ecoregions, species list**
- **SDMs for multiple taxa at same time**

Webinar organization

1. **Exploring Concepts:**

   Fundamental vs Realized Niche

2. **Demonstrations:**

   Applications

3. **Exercises:**

   Input occurrence records, ecoregions, species lists; SDMs

4. **Session Summary, Q&A and Discussion**

Species Distribution Modelling

BiotaPhy

SDM

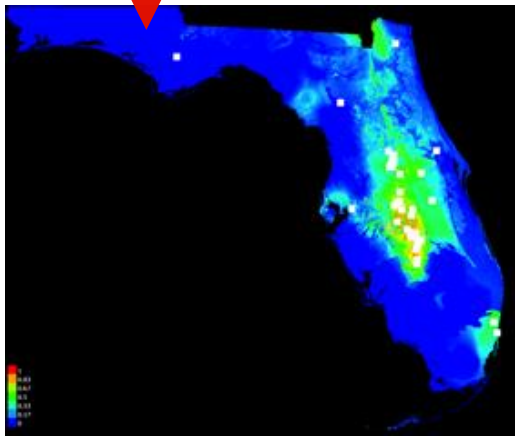
Species Distribution Modeling (SDM) is also known by several other names, including environmental niche modeling, ecological niche modeling, and habitat modeling. SDM refers to the process of creating mathematical formulas (models) to predict the geographic distribution of species based on where they have been found and the environmental conditions in those locations.

Species Distribution Model

A species distribution model (SDM) is an estimation of potential habitat for a particular species.

Prunus geniculata **(scrub plum)**

SDM for *Prunus geniculata*



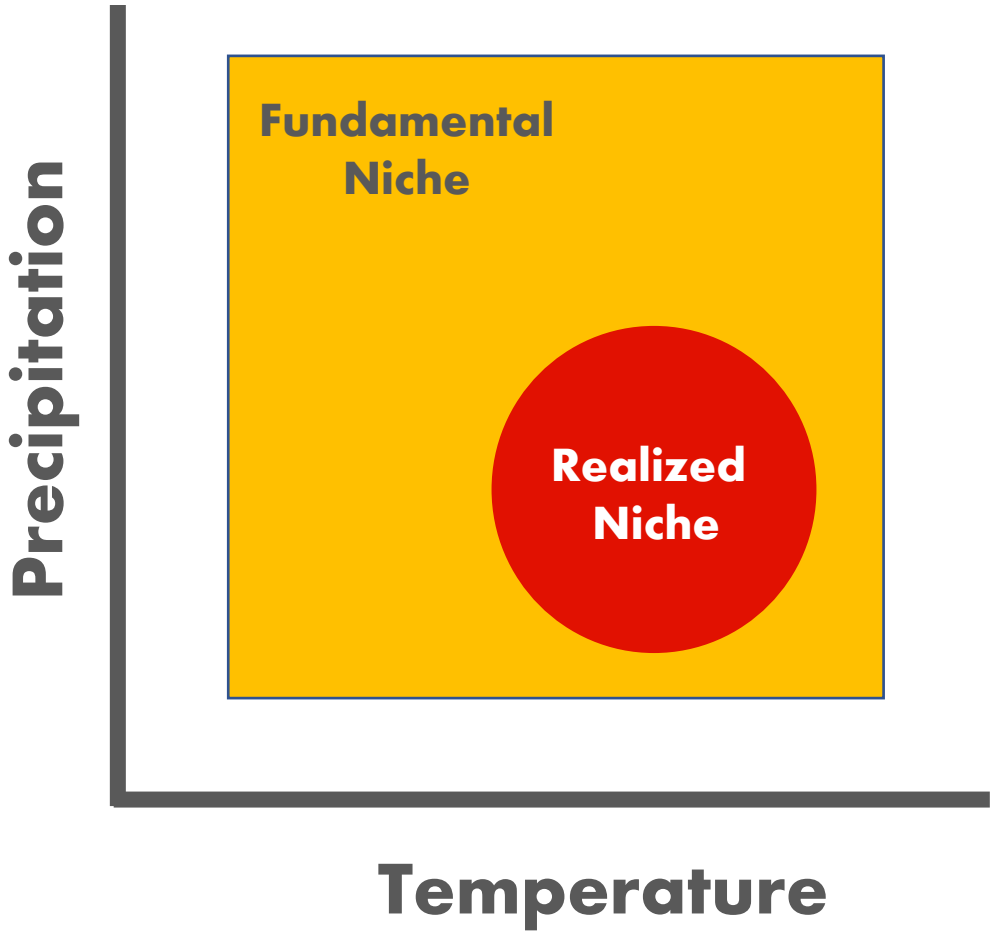
Exploring Concepts: Ecological Niche

Fundamental Niche

- abiotic conditions a species could potentially occupy in the absence of biotic interactions

Realized Niche

- abiotic conditions that a species can occupy with the presence of biotic interactions

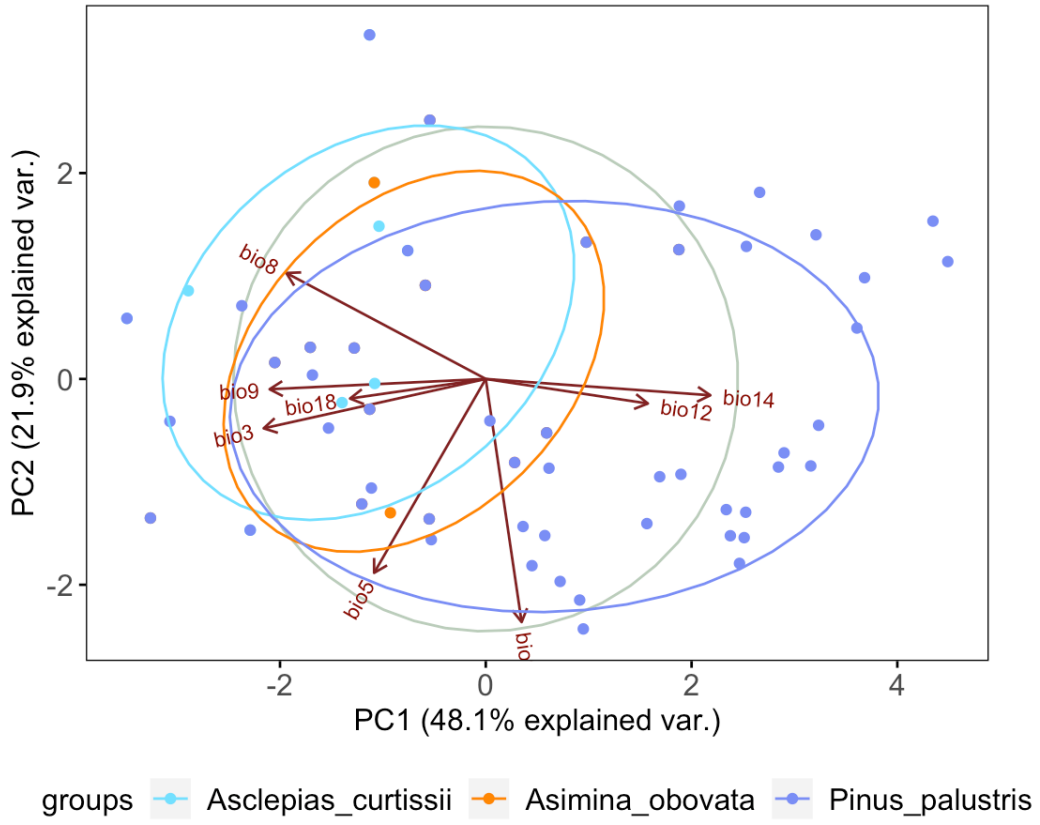


Exploring Concepts: Ecological Niche

Fundamental Niche

- abiotic conditions a species could potentially occupy in the absence of biotic interactions

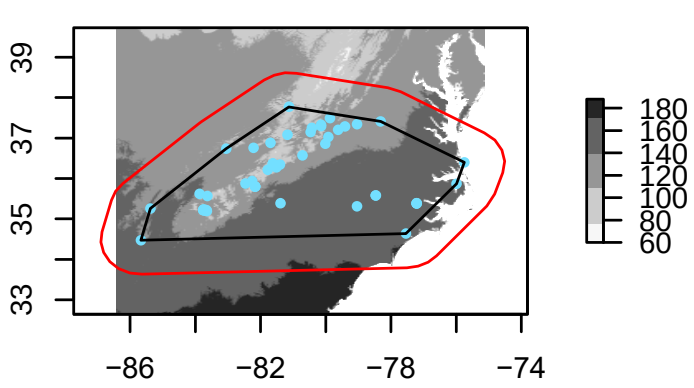
"it is defined in multidimensional ecological space (MacArthur 1972)." – Peterson 2001



Exploring Concepts: MaxEnt

MaxEnt

- uses the principle of maximum entropy on presence-only data to predict the species' potential geographic distribution (or niche)



Exploring Concepts: Applications

- Estimate current distribution
- Predict future distributions
- Infer past distributions
- Use as foundation for phylogenetic diversity studies
- Niche of polyploid compared to its diploid progenitor(s)
- Invasives--projections

Allen, J., Folk, R.A., P.S. Soltis, D.E. Soltis, R.P. Guralnick. 2019. Biodiversity synthesis across the green branches of the tree of life. *Nature Plants* 5: 11-13.

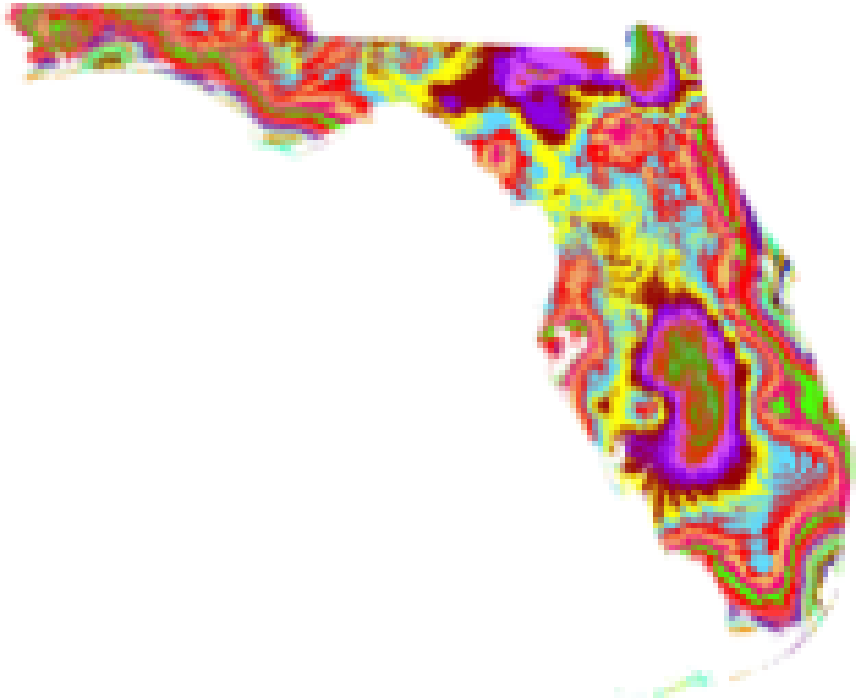
Exploring Concepts: Applications

- **Estimate current distribution**
- Predict future distributions
- Infer past distributions
- Use as foundation for phylogenetic diversity studies
- Niche of polyploid compared to its diploid progenitor(s)
- Invasives--projections

Allen, J., Folk, R.A., P.S. Soltis, D.E. Soltis, R.P. Guralnick. 2019. Biodiversity synthesis across the green branches of the tree of life. *Nature Plants* 5: 11-13.

Florida Plant Diversity in a Changing Climate!

Present and Future

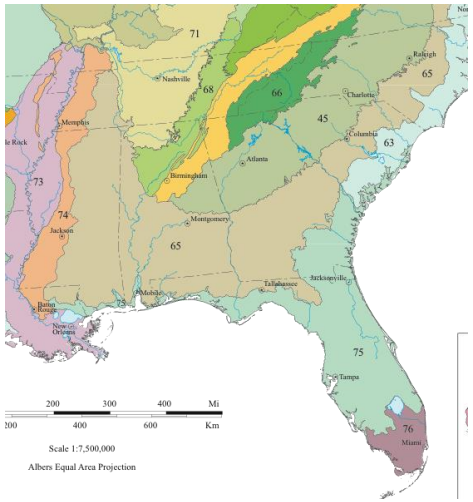


Julie Allen, Charlotte Germain-Aubrey, Rob Guralnick

Allen, Germain-Aubrey et al. 2019. *iScience* 11: 57–70 https://doi.org/10.1016/j.isci.2018.12.002

Florida Plant Diversity in a Changing Climate!

BiotaPhy



Present and Future

Florida Ecoregions
~4,100 vascular plant species

- Southeastern Plain
- Southern Coastal Plain
- Southern Florida Coastal Plain

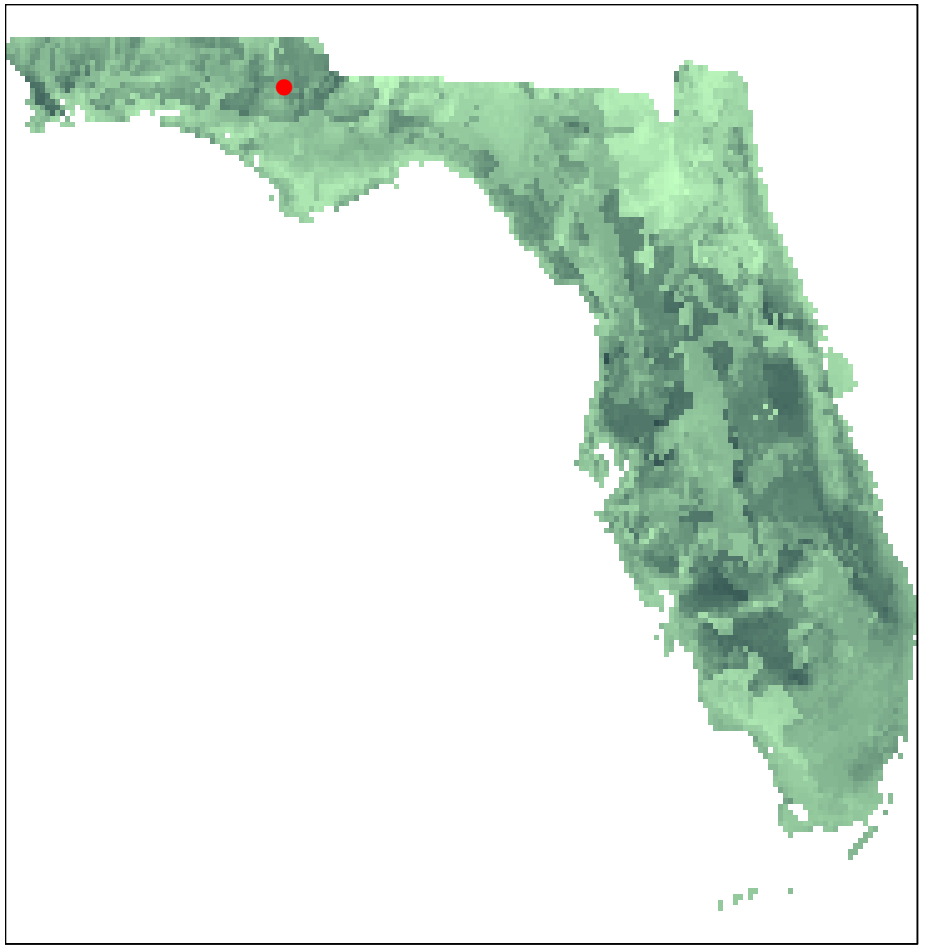
Modeling the Distribution of Species

- Location information from herbarium labels
- Environmental data:
  - Temperature, precipitation, soil, etc.
- Software to model the range of each species
- For Florida plants:
  - ~1500 plant species

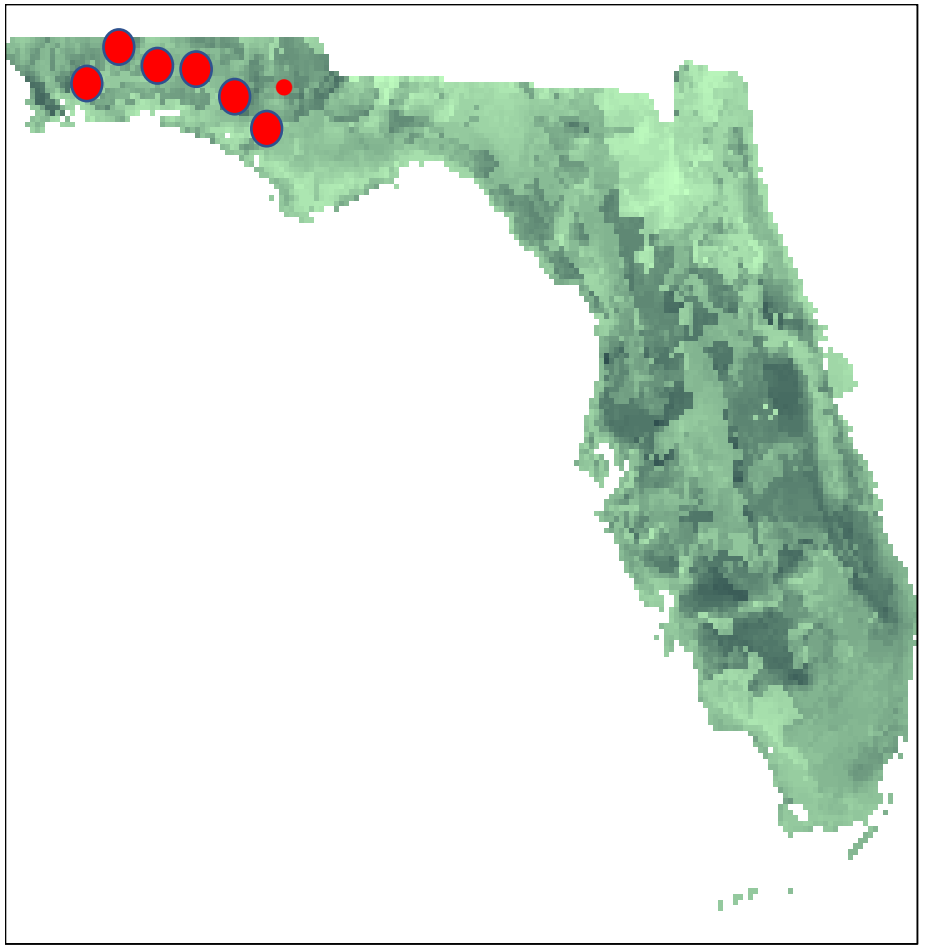
    (of 4100 species)

Modeling the Distribution of Species: Present

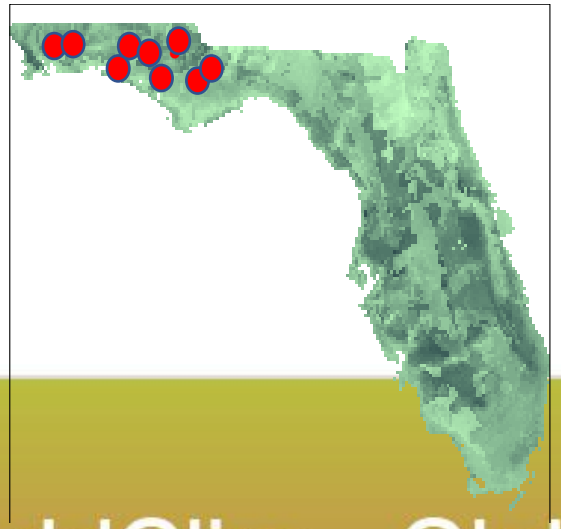
Modeling the Distribution of Species: Present



Modeling the Distribution of Species: Present



Modeling the Distribution of Species: Present

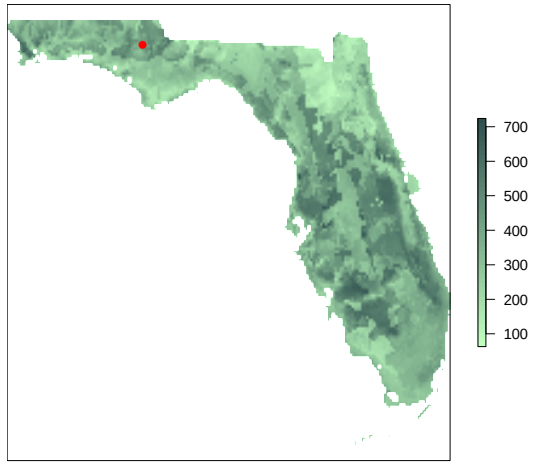


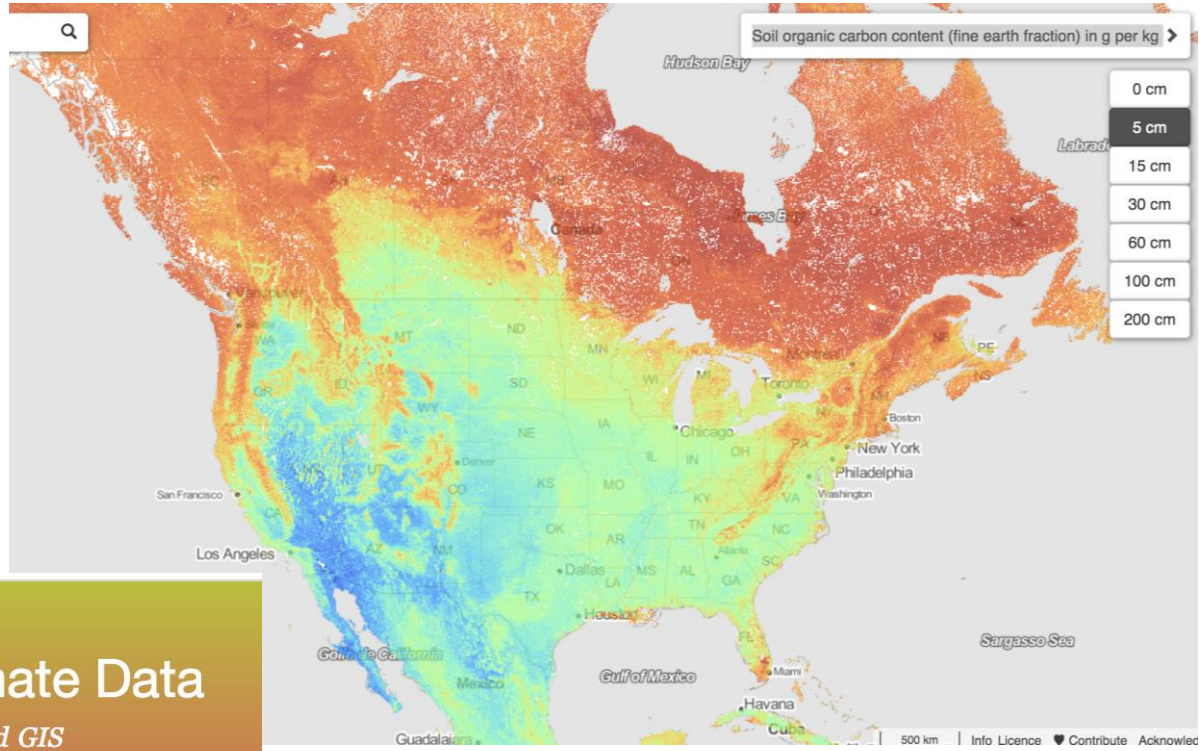
WorldClim - Global Climate Data

Free climate data for ecological modeling and GIS

19 **Bioclimatic variables**
temperature
rainfall

Modeling the Distribution of Species: Present

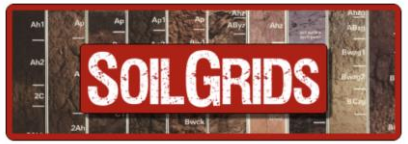




WorldClim - Global Climate Data

Free climate data for ecological modeling and GIS

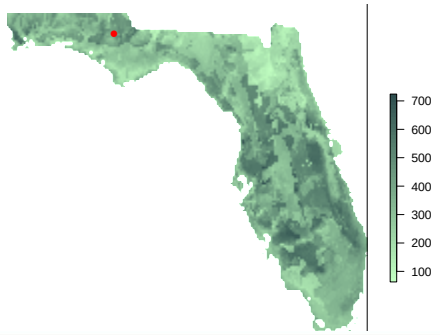
Global maps of soil properties at multiple depths



b cycle background maps a geolocate m open layer menu
i open SoilInfo *(works only when location is selected)*

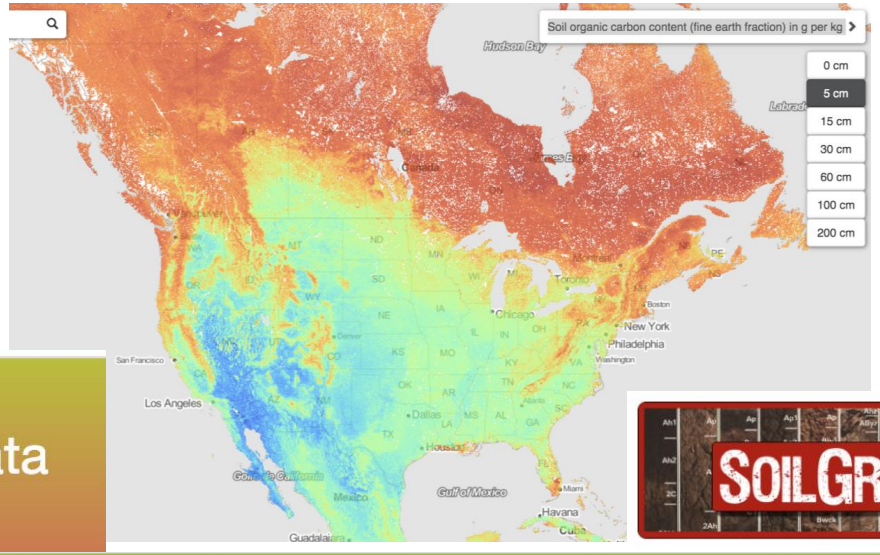
Powered by ISRIC World Soil Information

Modeling the Distribution of Species: Present



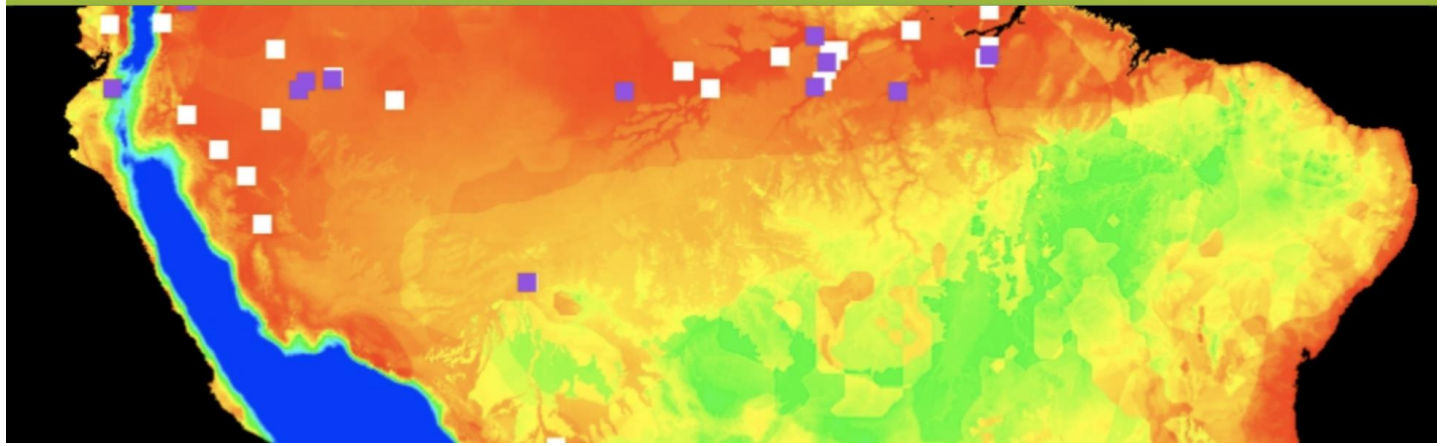
WorldClim - Global Climate Data

Free climate data for ecological modeling and GIS



SOILGRIDS

Maxent software for modeling species niches and distributions



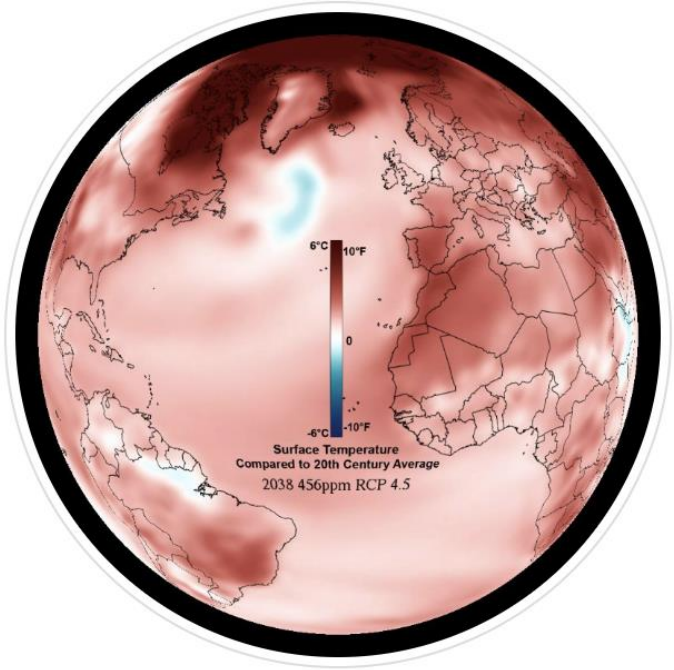
Exploring Concepts: Applications

- Estimate current distribution
- **Predict future distributions**
- Infer past distributions
- Use as foundation for phylogenetic diversity studies
- Niche of polyploid compared to its diploid progenitor(s)
- Invasives--projections

Allen, J., Folk, R.A., P.S. Soltis, D.E. Soltis, R.P. Guralnick. 2019. Biodiversity synthesis across the green branches of the tree of life. *Nature Plants* 5: 11-13.

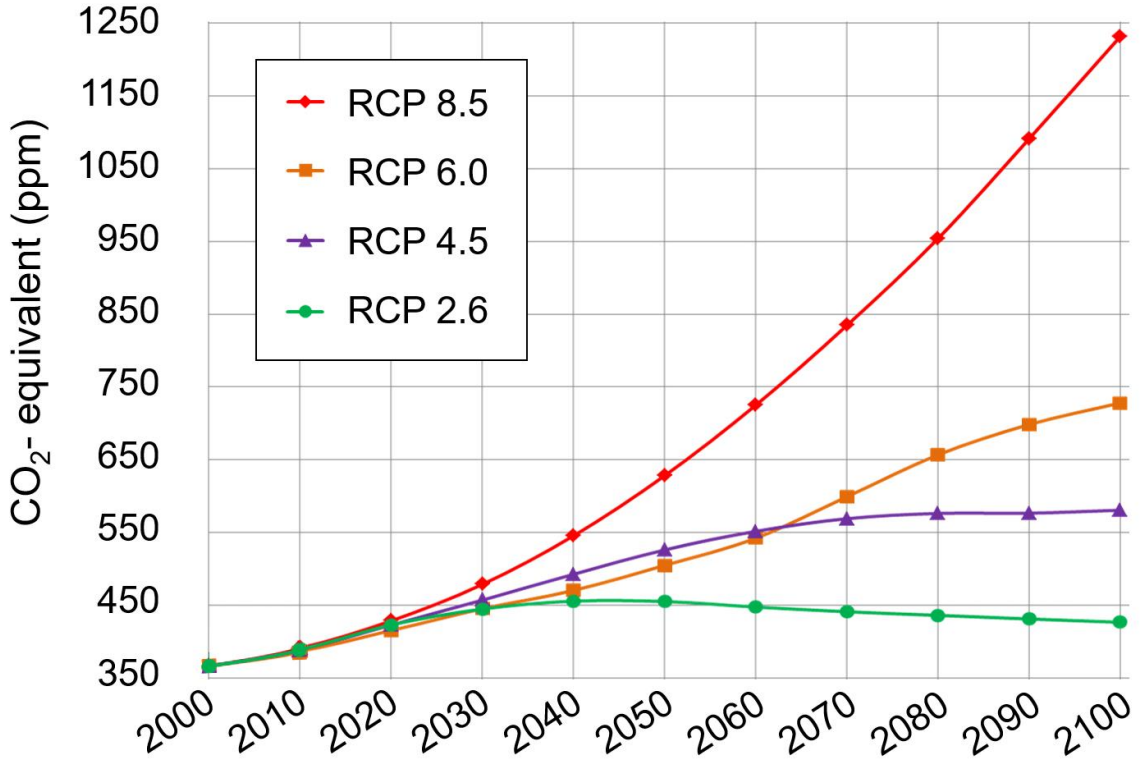
Modeling the Distribution of Species: Future

Representative Concentration Pathways: Models for Future Temperatures



IPCC AR5 Greenhouse Gas Concentration Pathways

Representative Concentration Pathways (RCPs) from the fifth Assessment Report by the International Panel on Climate Change



Continues to rise – 3.7 C

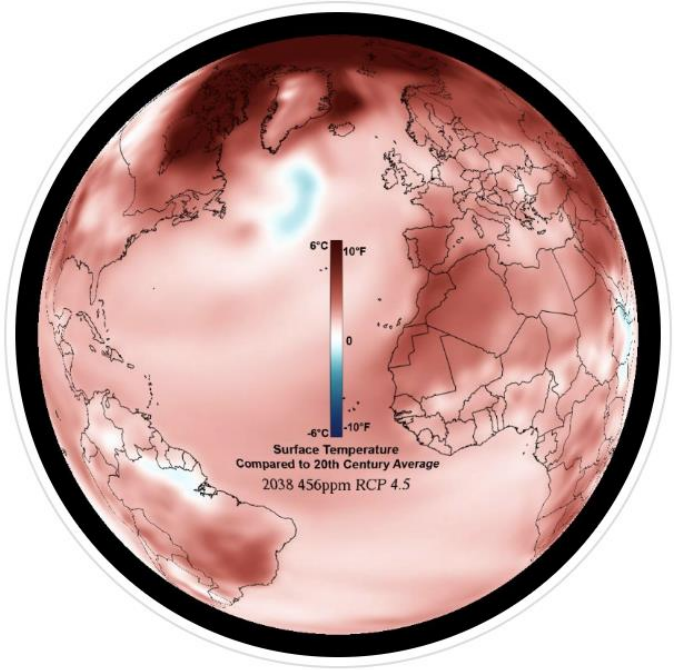
Peak in 2060 – 2.2 C

Peak in 2040 – 1.8 C

Peak in 2020 – 1 C

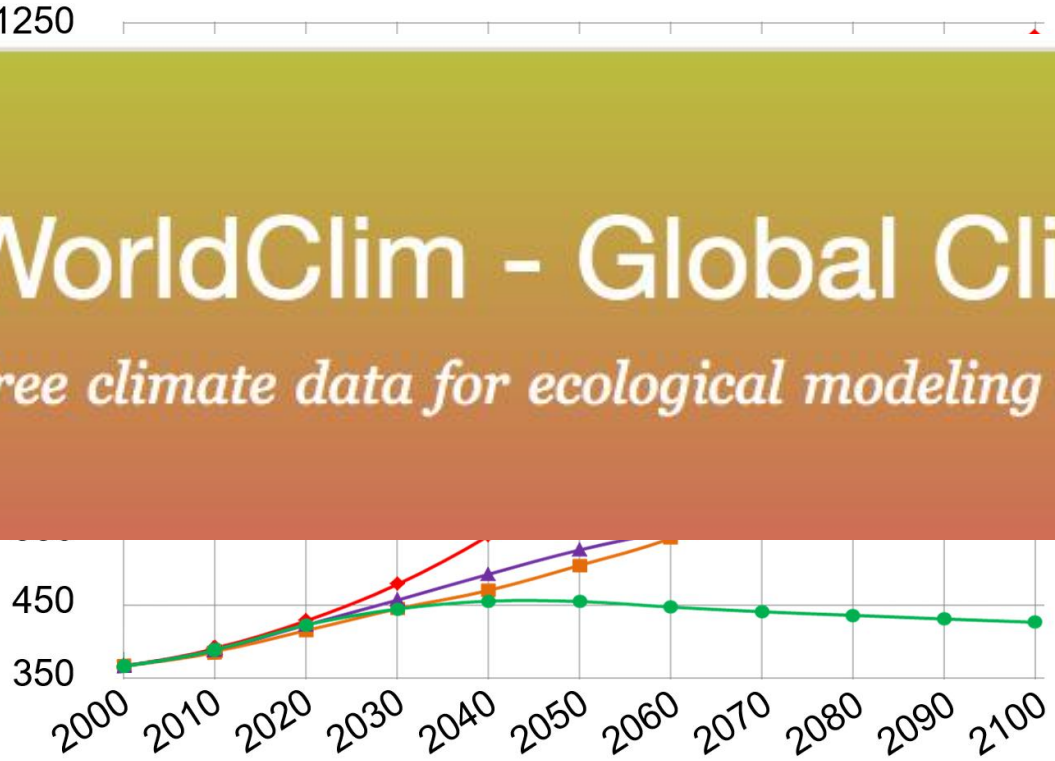
Modeling the Distribution of Species: Future

Representative Concentration Pathways: Models for Future Temperatures



IPCC AR5 Greenhouse Gas Concentration Pathways

Representative Concentration Pathways (RCPs) from the fifth Assessment Report by the International Panel on Climate Change



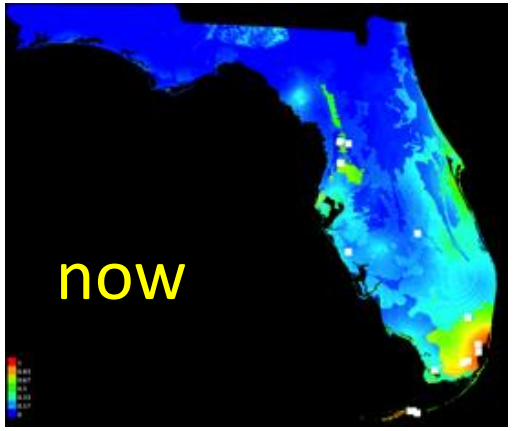
WorldClim - Global Climate Data

Free climate data for ecological modeling and GIS

Response to Climate Change

***Abildgaardia ovata* (flatspike sedge)**

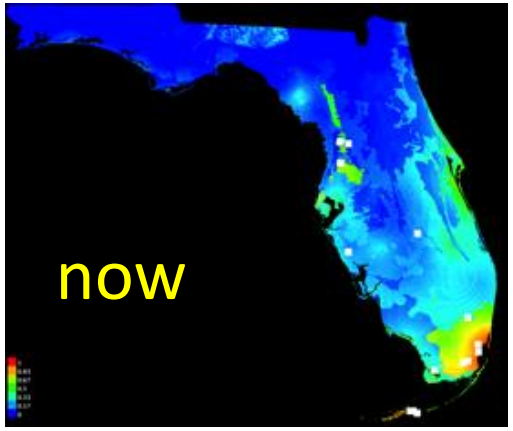
Models for Present and Future!

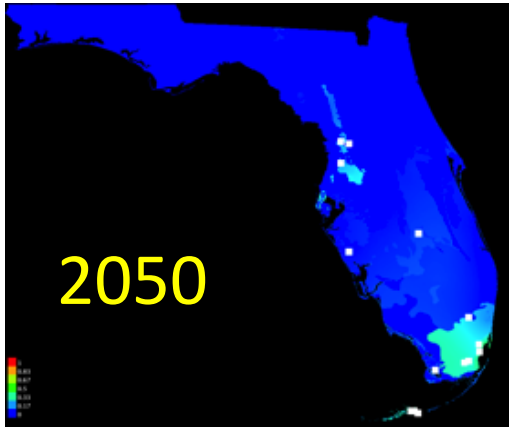


Response to Climate Change

***Abildgaardia ovata* (flatspike sedge)**

Models for Present and Future!

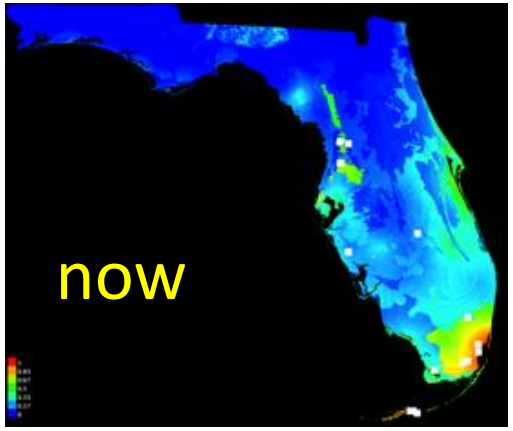


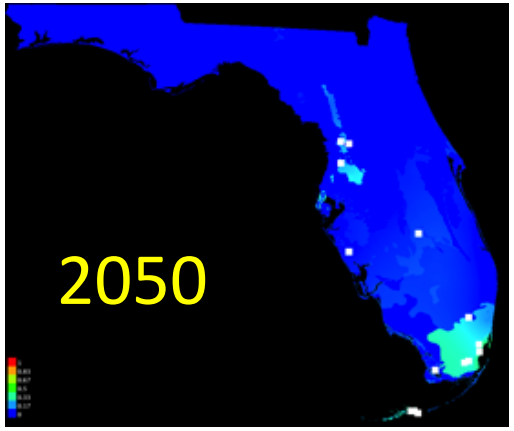


Response to Climate Change

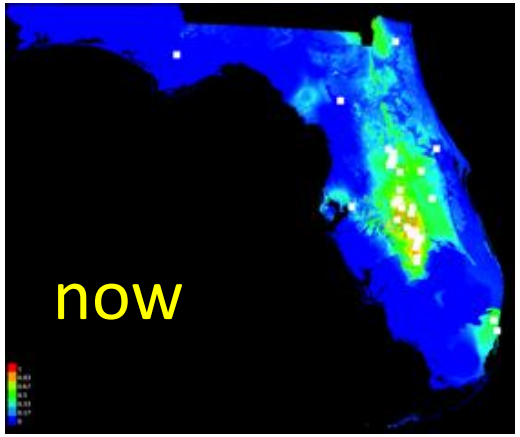
Models for Present and Future!

***Abildgaardia ovata* (flatspike sedge)**





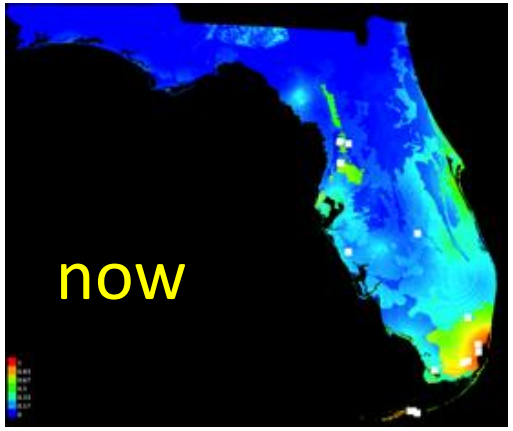
***Prunus geniculata* (scrub plum)**

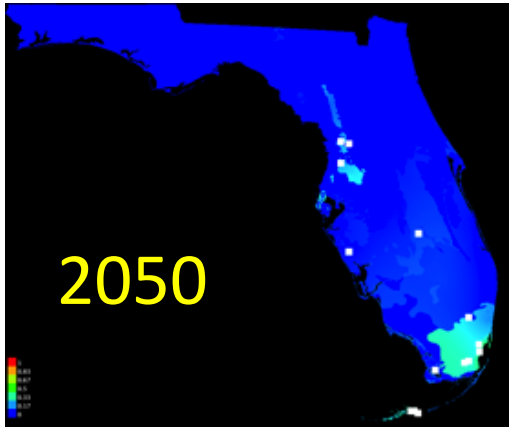


Response to Climate Change

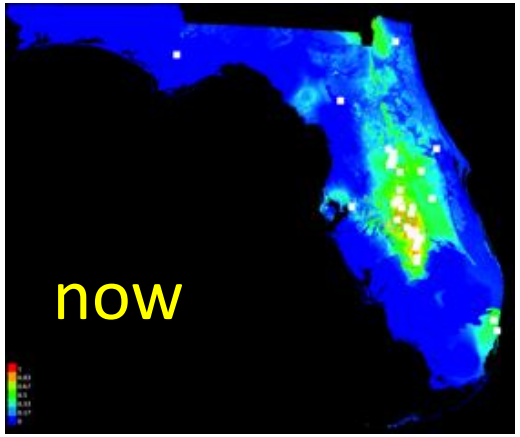
Models for Present and Future!

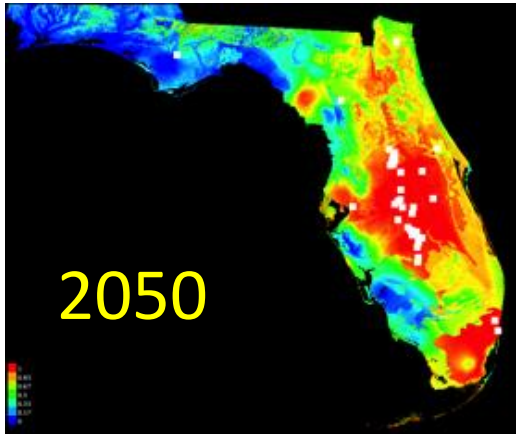
Abildgaardia ovata (flatspike sedge)





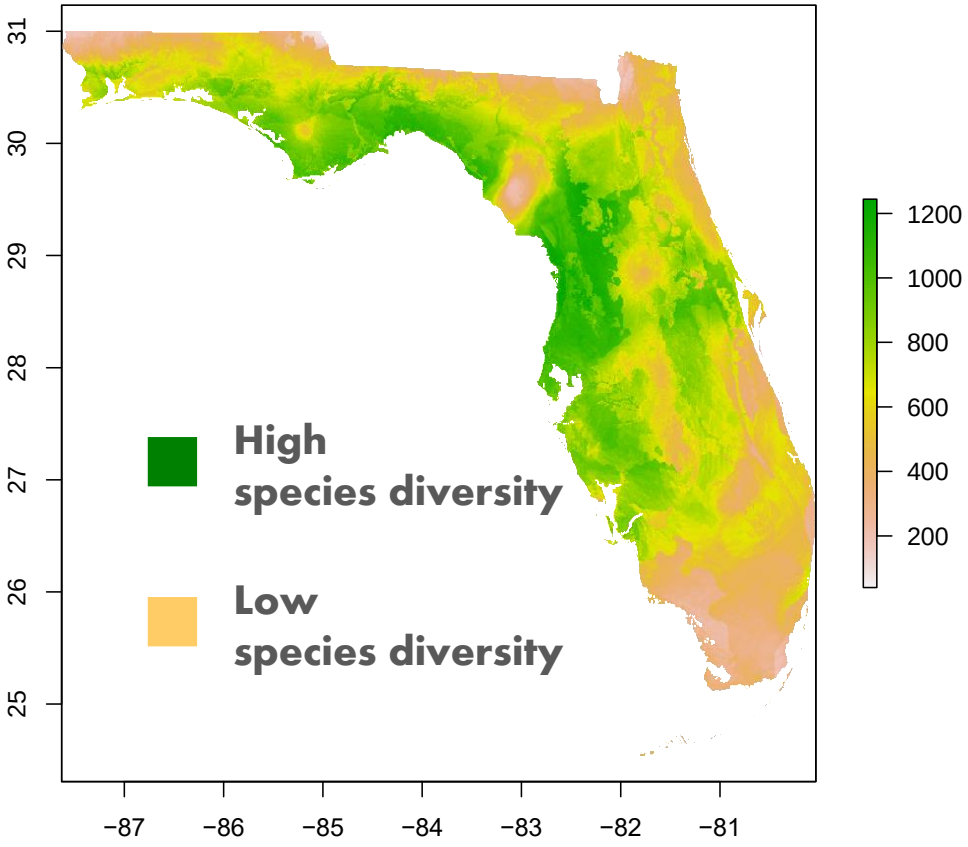
Prunus geniculata (scrub plum)





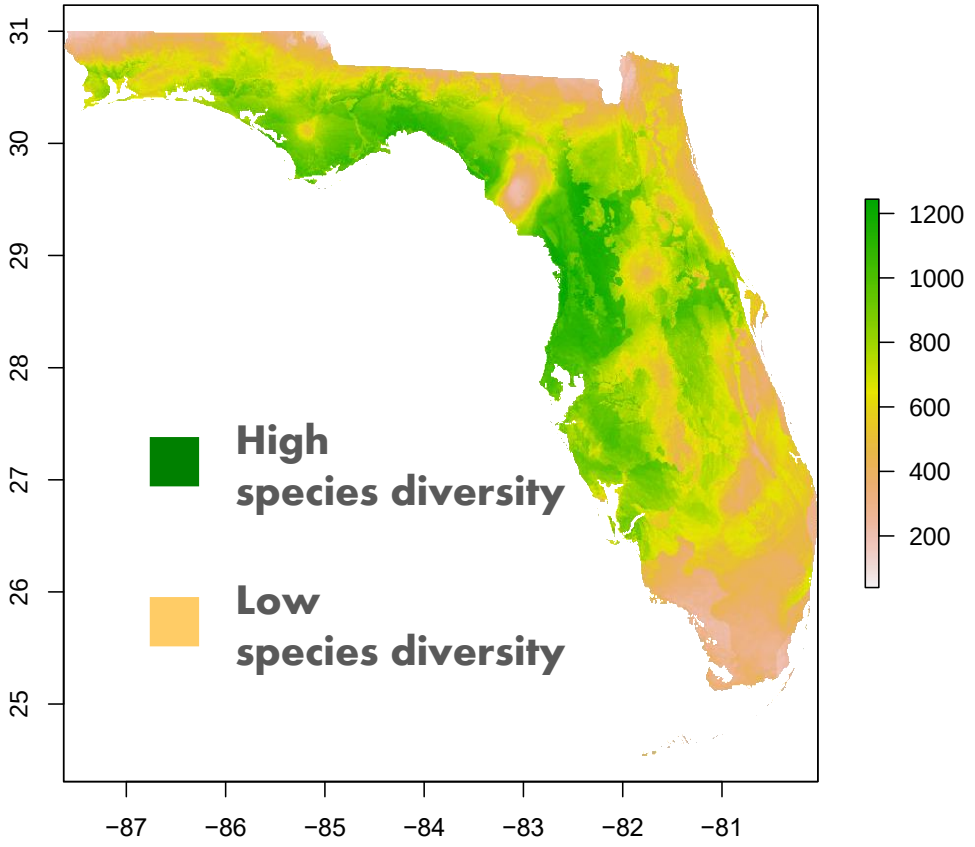
Florida Plant Diversity: Now & 2050

Present

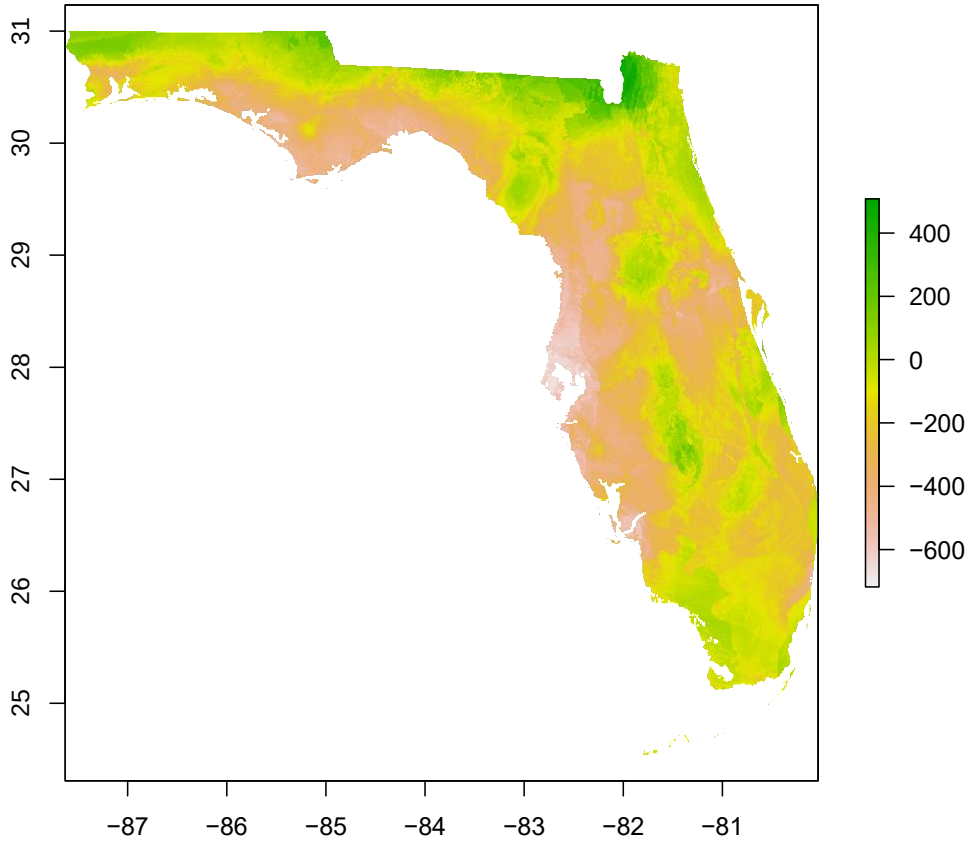


Florida Plant Diversity: Now & 2050

Present



spp 2050 - #spp now

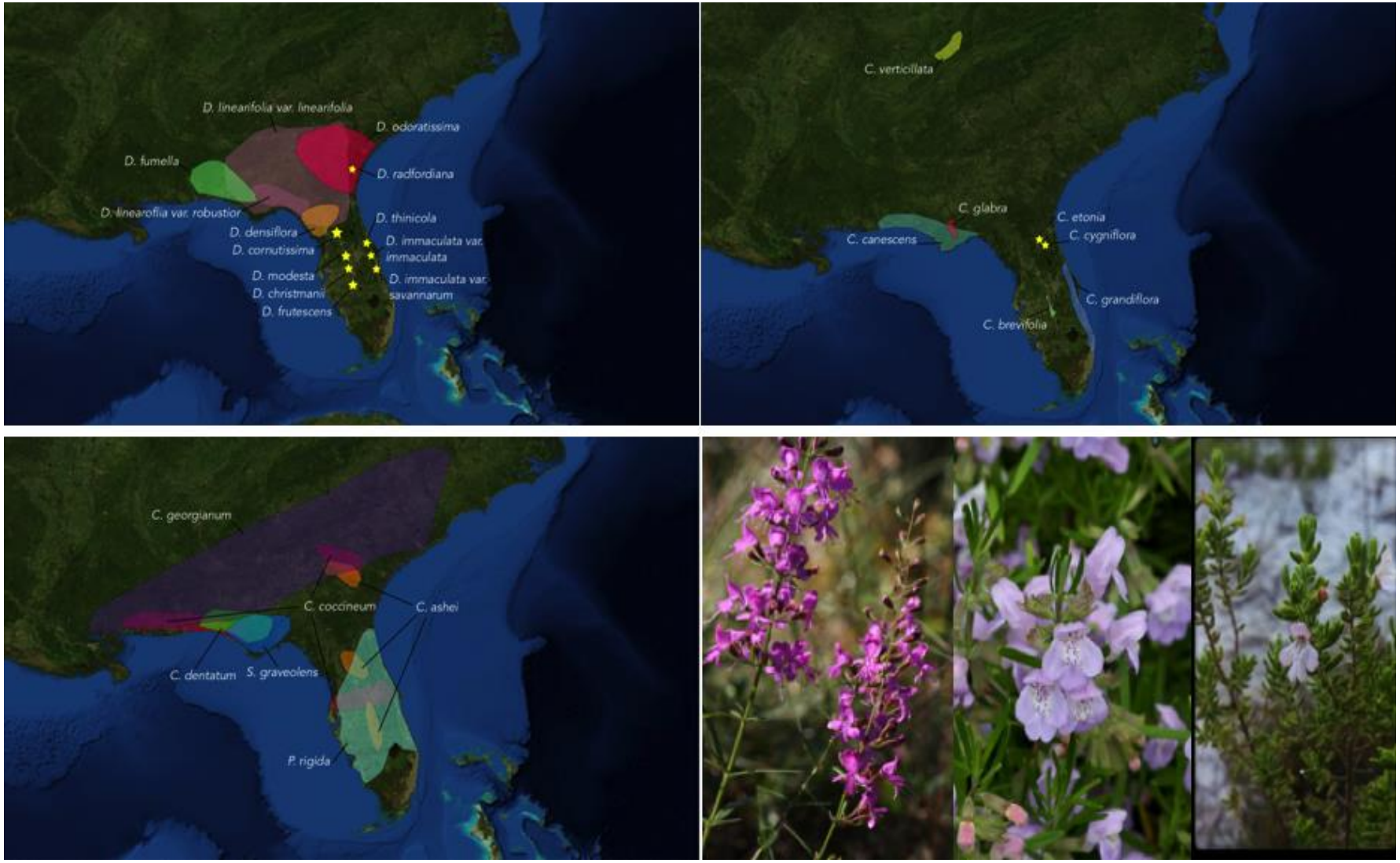


Exploring Concepts: Applications

- Estimate current distribution
- Predict future distributions
- **Infer past distributions**
- Use as foundation for phylogenetic diversity studies
- Niche of polyploid compared to its diploid progenitor(s)
- Invasives--projections

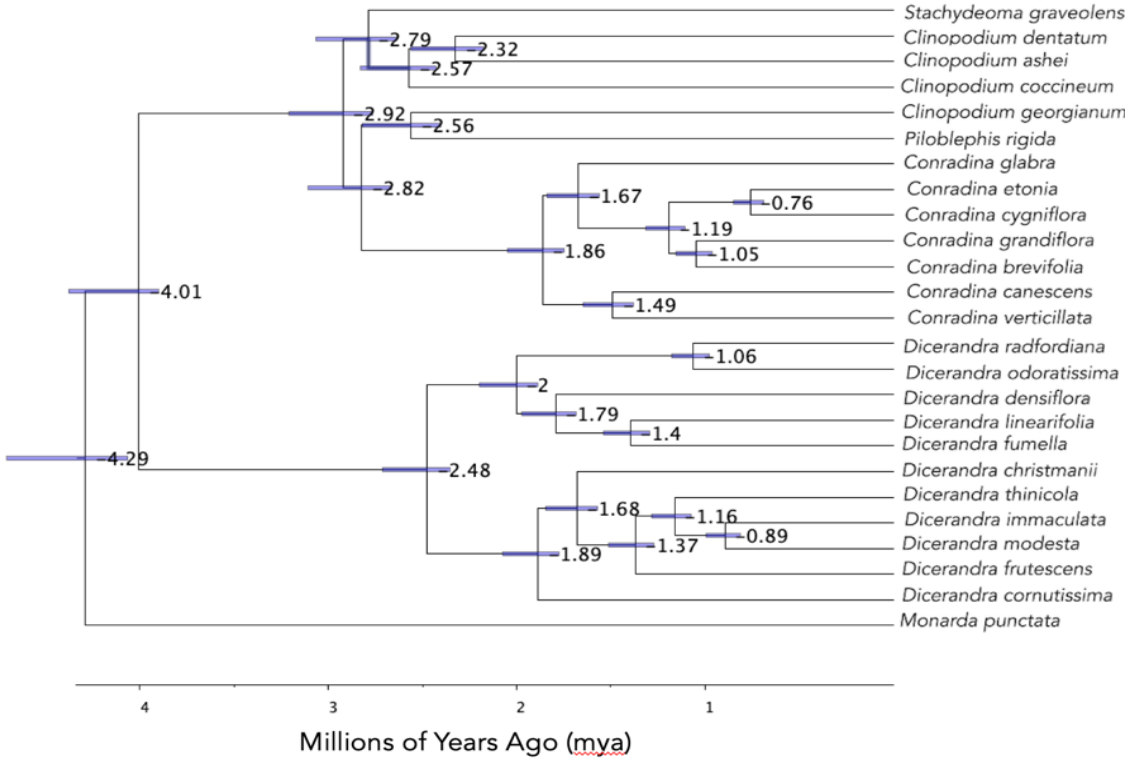
Allen, J., Folk, R.A., P.S. Soltis, D.E. Soltis, R.P. Guralnick. 2019. Biodiversity synthesis across the green branches of the tree of life. *Nature Plants* 5: 11-13.

Past Distribution: Scrub Mint Clade



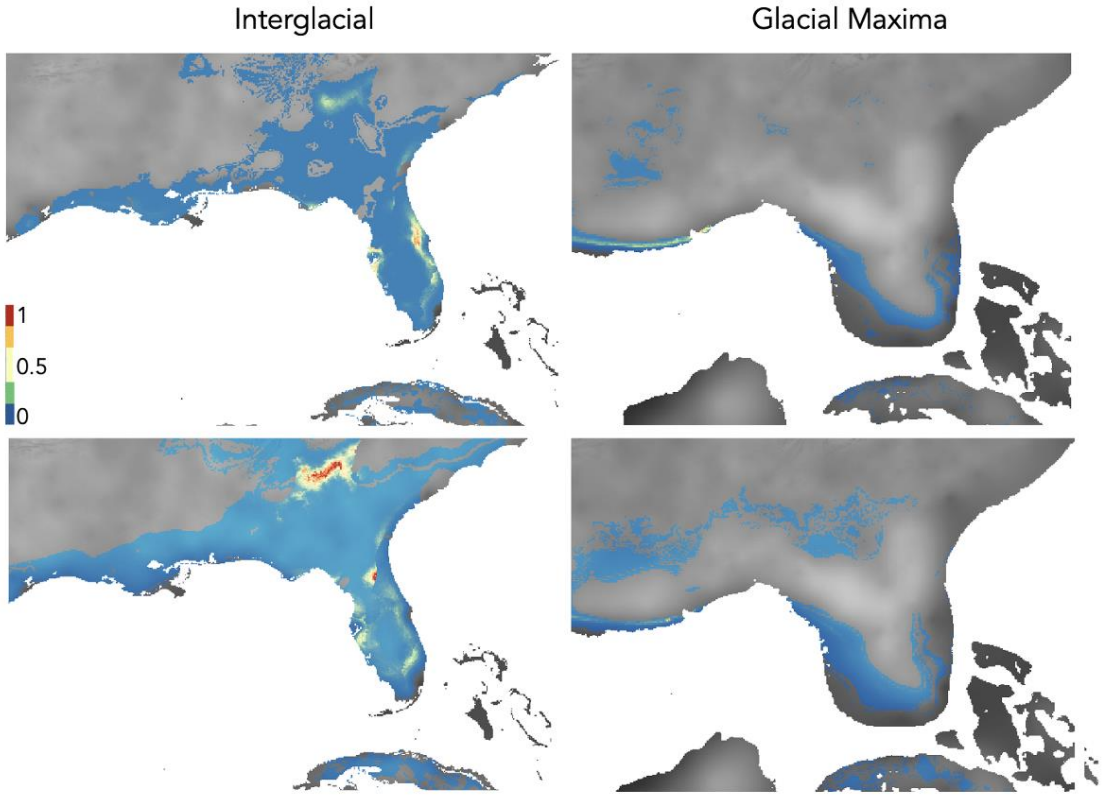
Naranjo et al. 2022.
J. Biogeography

Past Distribution: Scrub Mint Clade



MRCA of SMC

MRCA of *Dicerandra*



PaleoClim: paleoclim.org

Naranjo et al. 2022. *J. Biogeography*

Folk et al. 2018. *American Naturalist*

Exploring Concepts: Applications

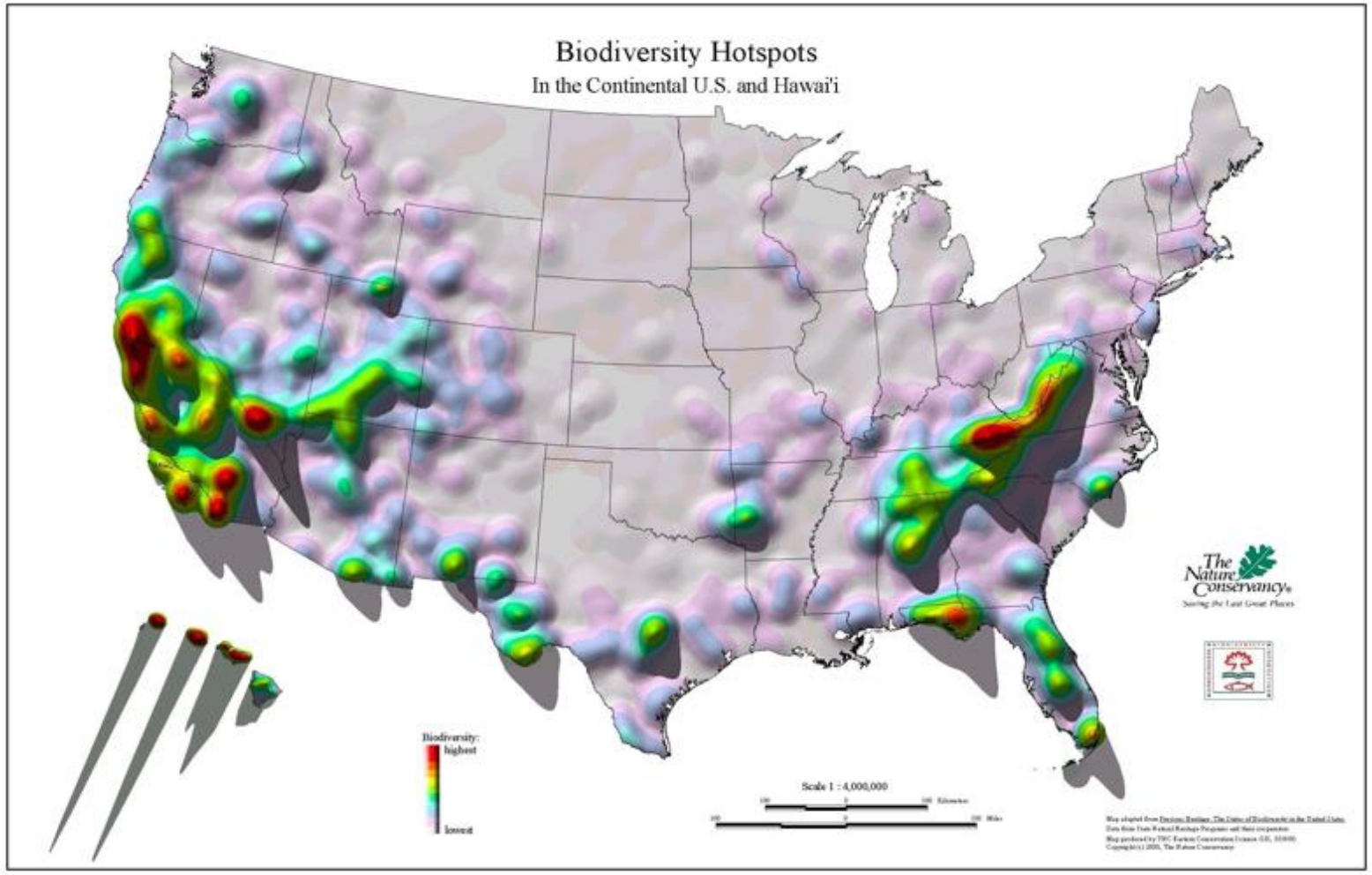
- Estimate current distribution
- Predict future distributions
- Infer past distributions
- **Use as foundation for phylogenetic diversity studies**
- Niche of polyploid compared to its diploid progenitor(s)
- Invasives--projections

Allen, J., Folk, R.A., P.S. Soltis, D.E. Soltis, R.P. Guralnick. 2019. Biodiversity synthesis across the green branches of the tree of life. *Nature Plants* 5: 11-13.

Exploring Concepts: Applications

BiotaPhy

What to Protect?

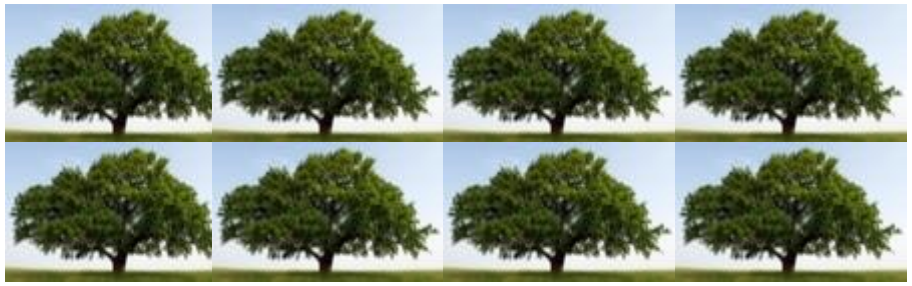


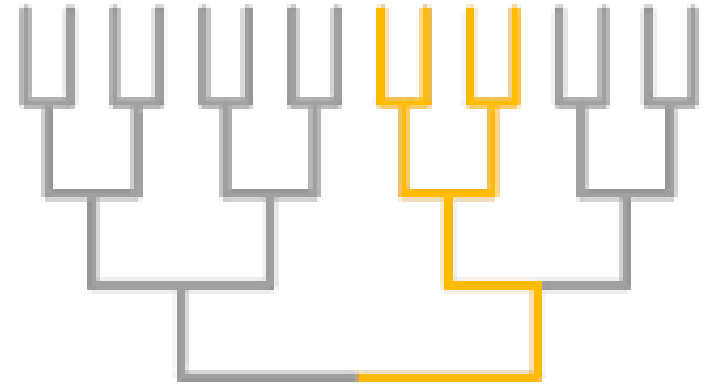
Phylogenetic Diversity

BiotaPhy

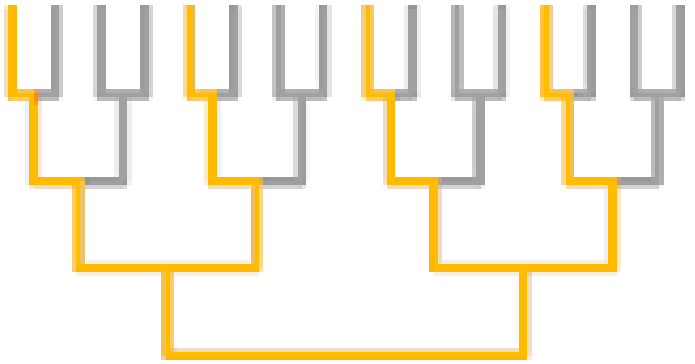
How much of the Tree of Life is present in a geographic area?

Oaks



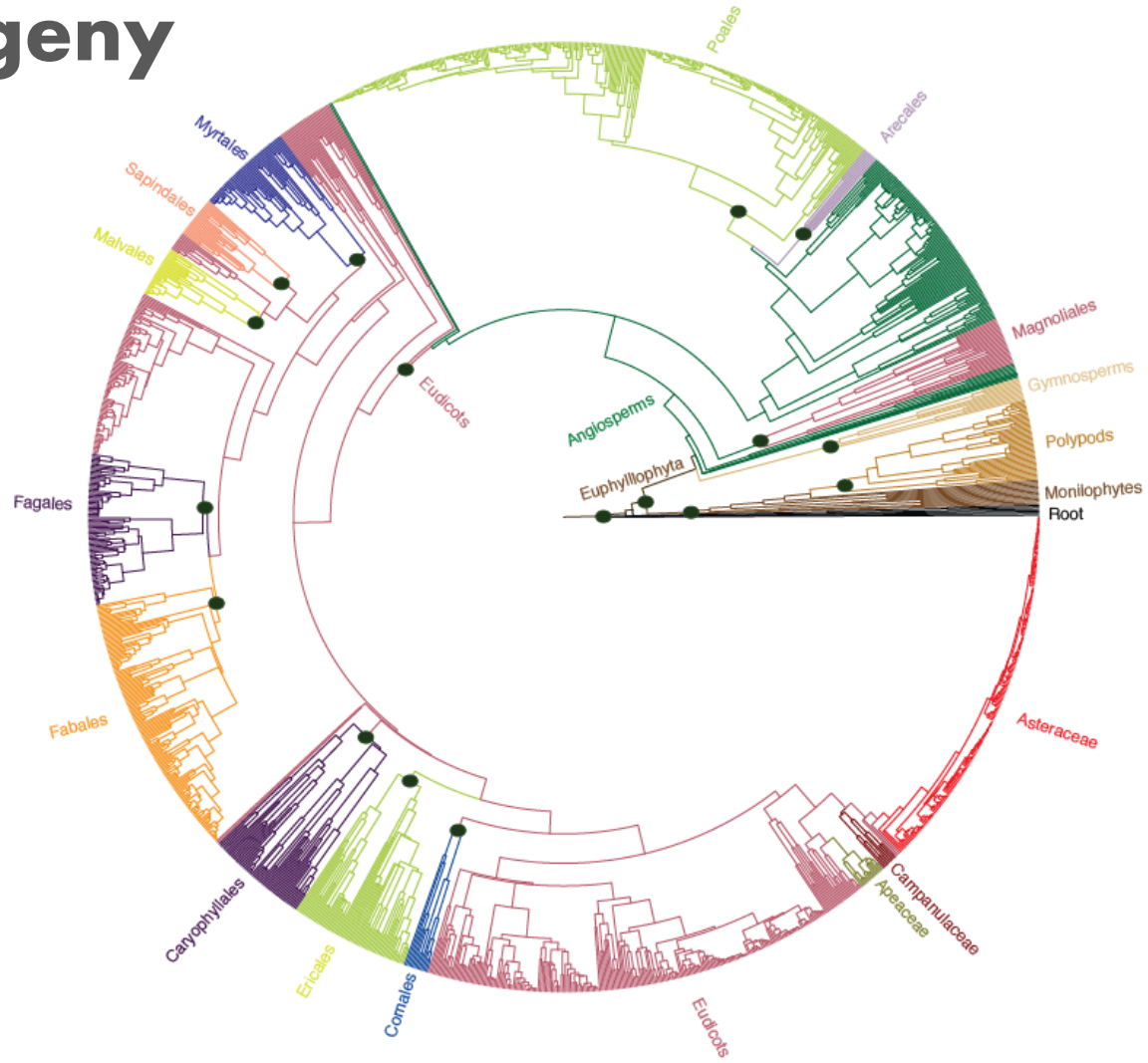


Vs.



Phylogenetic Diversity

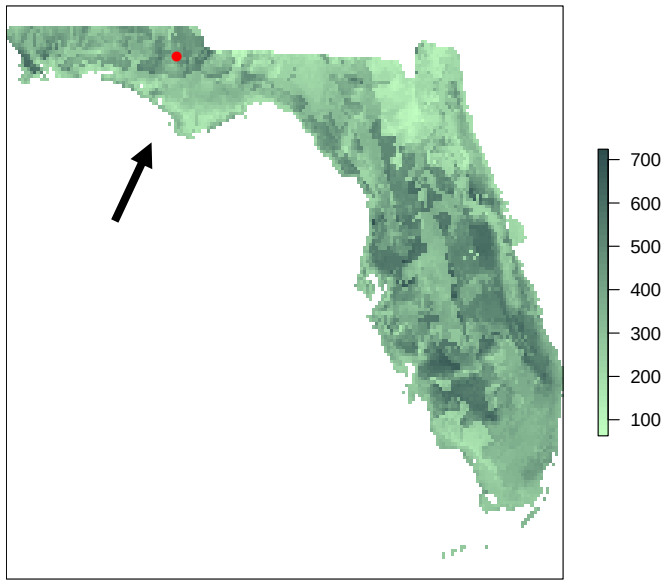
Florida Plant Phylogeny



Allen et al. 2019

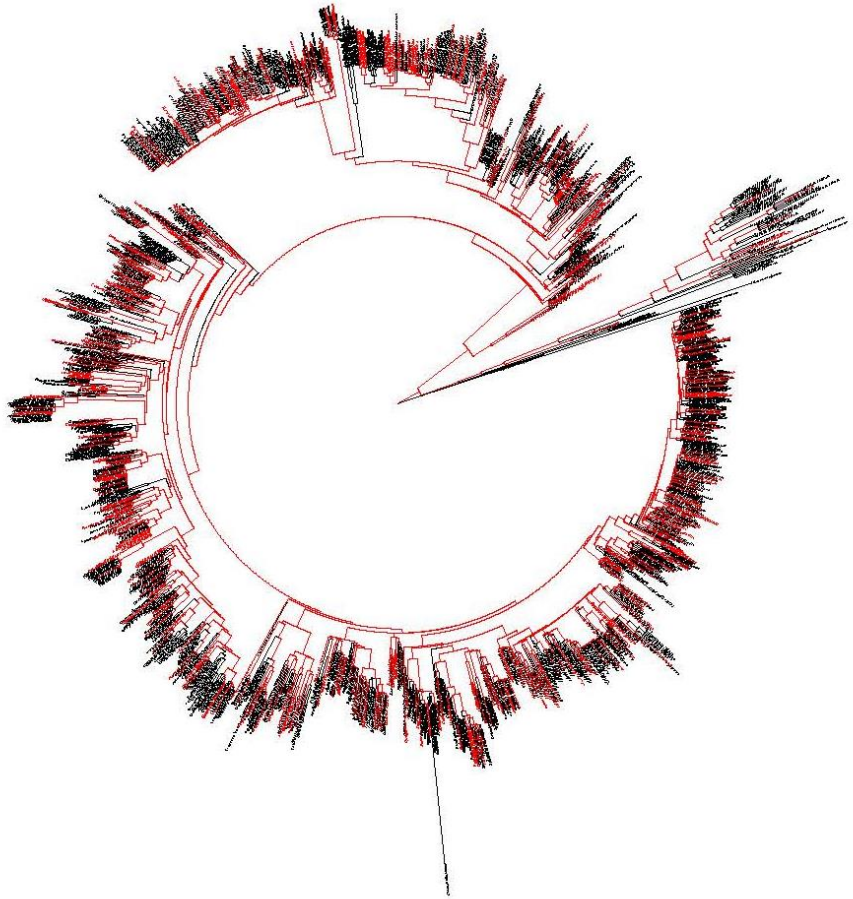
Computing Phylogenetic Diversity

How much of the Tree of Life is present in a geographic area?



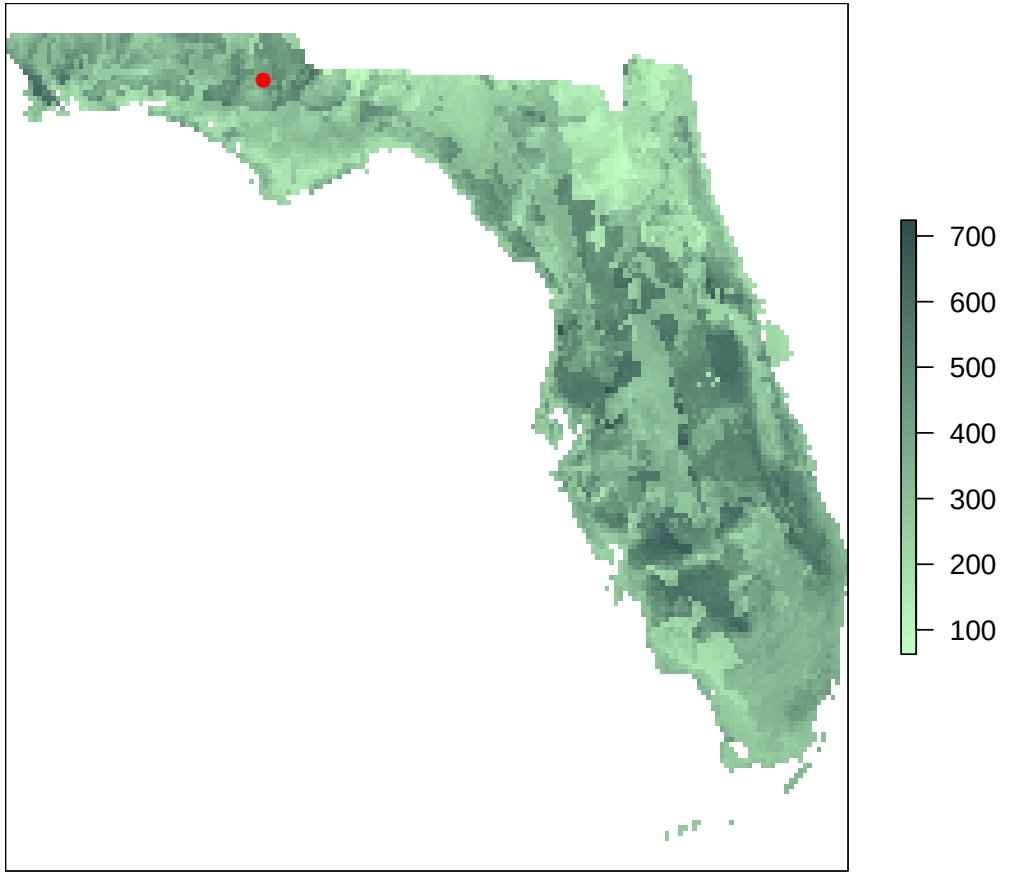
8,045 pixels/communities
16 km² per pixel

Species A
Species B
Species C
Species D
Species E
...



Computing Phylogenetic Diversity

How much of the Tree of Life is present in a geographic area?

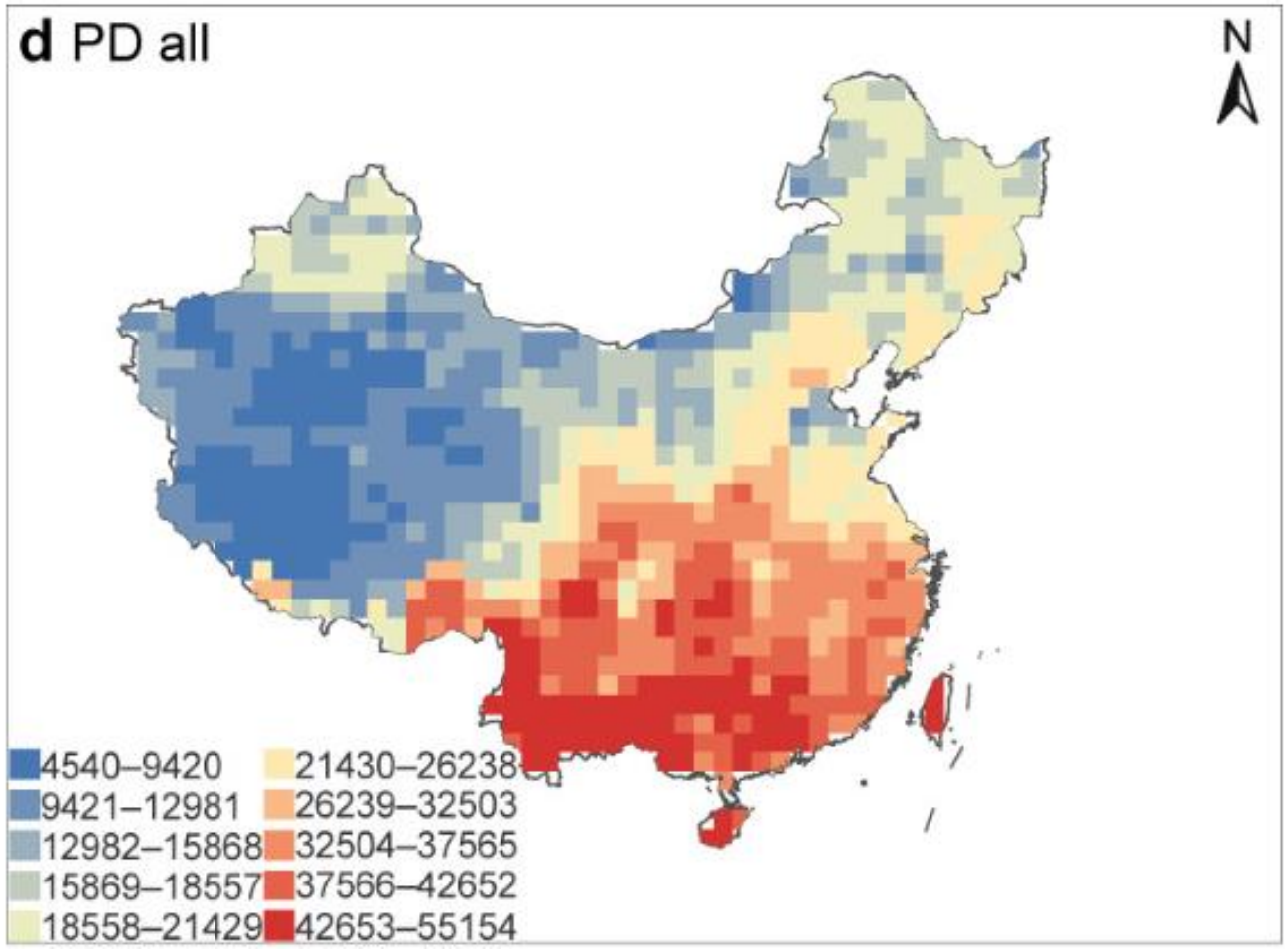


Phylogenetic Diversity: China

How much of the Tree of Life is present in a geographic area?

Low PD

High PD



Lu et al. 2018. *Nature*

Exploring Concepts: Applications

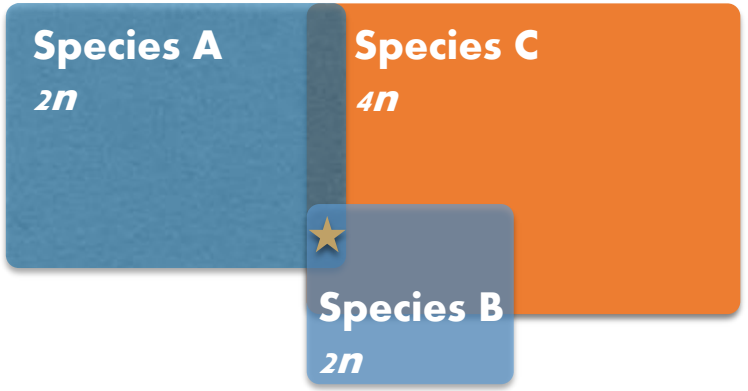
- Estimate current distribution
- Predict future distributions
- Infer past distributions
- Use as foundation for phylogenetic diversity studies
- **Niche of polyploid compared to its diploid progenitor(s)**
- Invasives--projections

Allen, J., Folk, R.A., P.S. Soltis, D.E. Soltis, R.P. Guralnick. 2019. Biodiversity synthesis across the green branches of the tree of life. *Nature Plants* 5: 11-13.

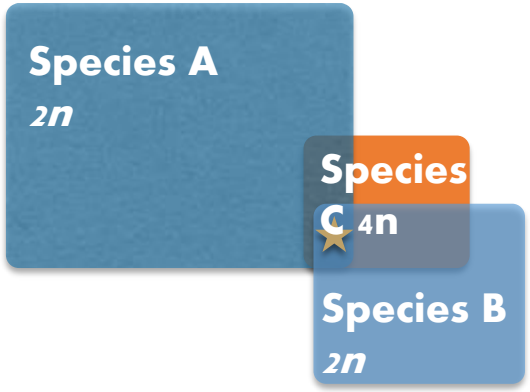
Niche Evolution in Allopolyploids

Niche Intermediacy

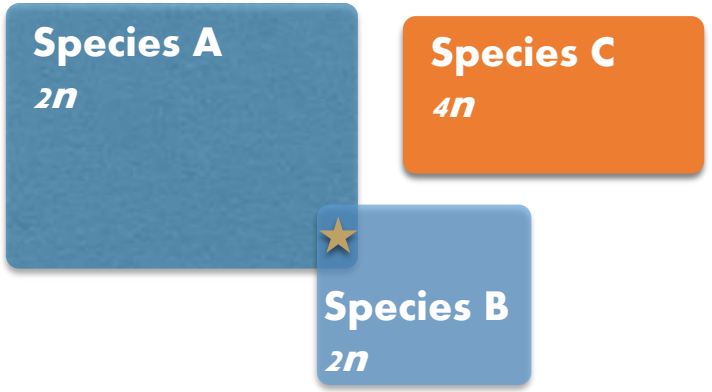
Niche Expansion



Niche Contraction



Niche Novelty

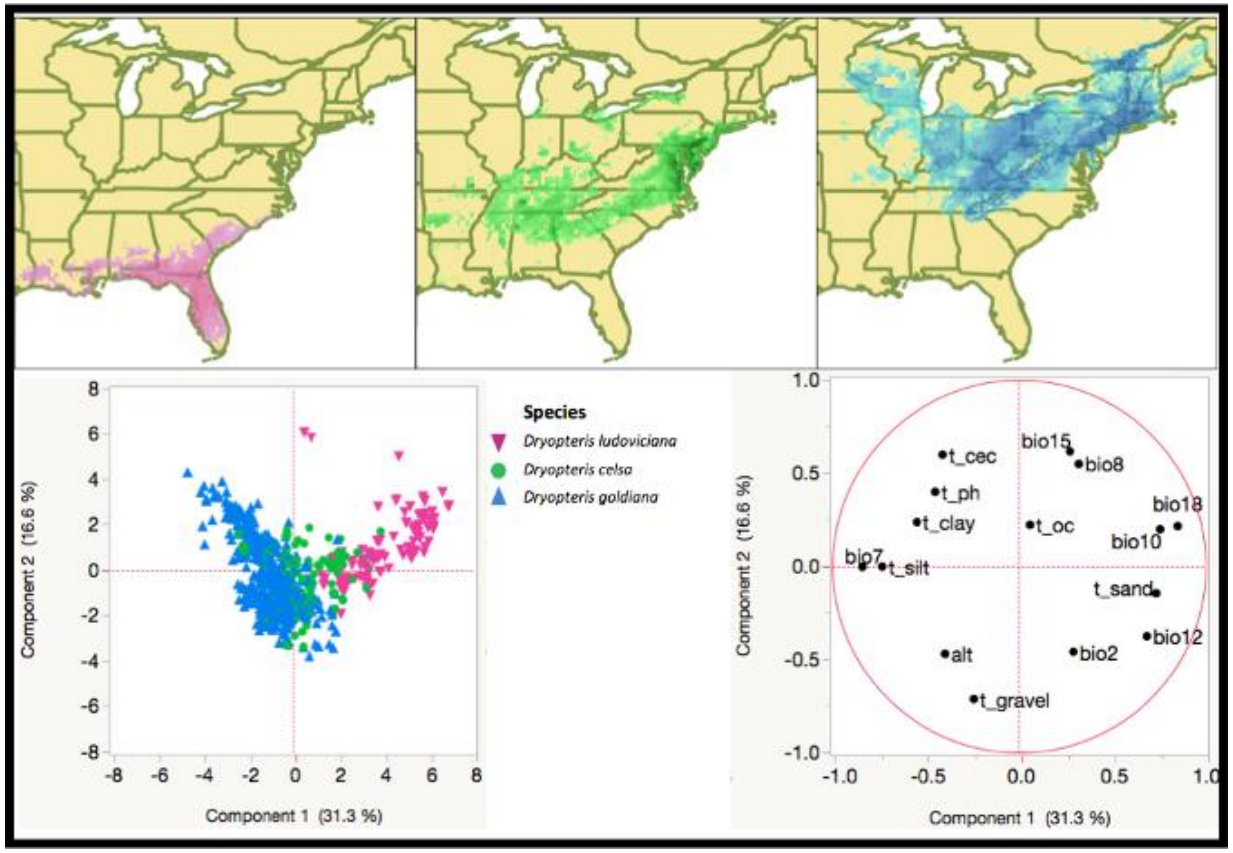


Blaine Marchant

Niche Evolution in Allopolyploids

Niche Evolution in Allopolyploids

Niche Intermediacy
Breadth: Parent < Polyploid < Parent
Overlap: Polyploid > 0.3



Dryopteris celsa

Nickrent, D.L. et al. 2006 onwards. *PhytoImages*. http://www.phytoimages.siu.edu

Niche Evolution in Allopolyploids

13 allopolyploids & parents

- Niche intermediacy: 8
- Niche contraction: 2
- Niche expansion: 2
- Niche novelty: 1

- More cases are needed!

Marchant et al. 2016. *New Phytologist*

Niche Evolution in Autopolyploids

RESEARCH ARTICLE

AMERICAN JOURNAL OF BOTANY

Niche divergence between diploid and autotetraploid *Tolmiea*[1]

Clayton J. Visger[2,3,5], Charlotte C. Germain-Aubrey[3], Maya Patel[3], Emily B. Sessa[2,4], Pamela S. Soltis[3,4], and Douglas E. Soltis[2-4]

Tolmiea menziesii

Galax urceolata

American Journal of Botany

RESEARCH ARTICLE

Climatic niche comparison among ploidal levels in the classic autopolyploid system, *Galax urceolata*

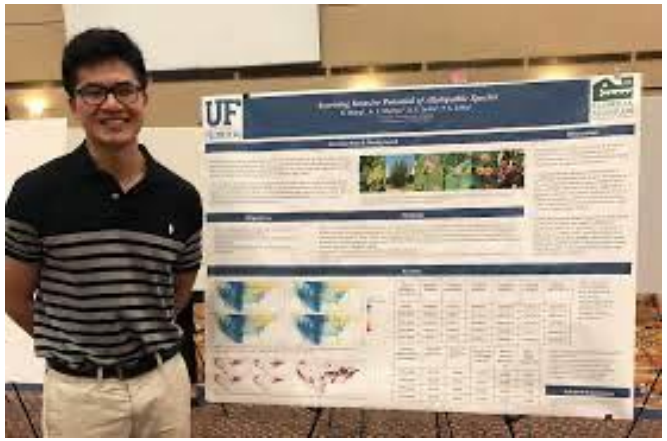
Michelle L. Gaynor[1,4], D. Blaine Marchant[2,3], Douglas E. Soltis[2,3], and Pamela S. Soltis[3]

Exploring Concepts: Applications

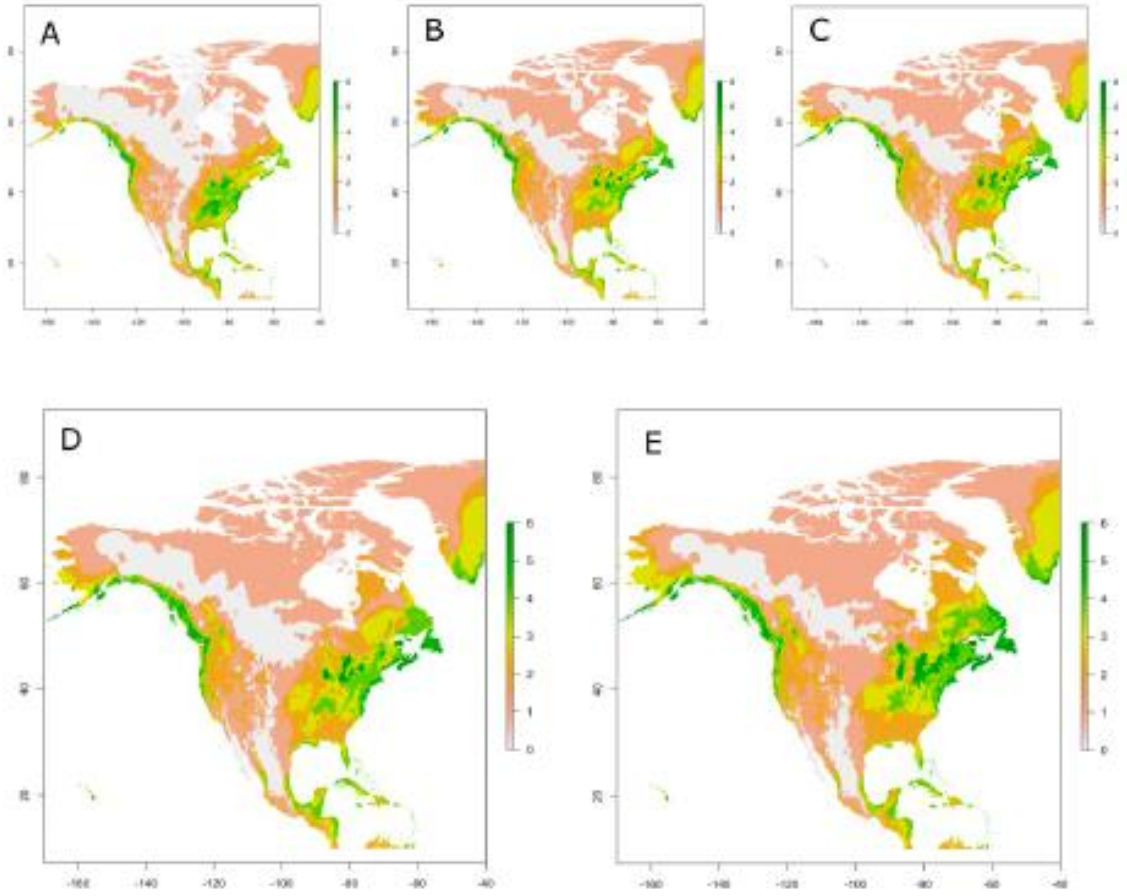
- Estimate current distribution
- Predict future distributions
- Infer past distributions
- Use as foundation for phylogenetic diversity studies
- Niche of polyploid compared to its diploid progenitor(s)
- Invasives--projections

Allen, J., Folk, R.A., P.S. Soltis, D.E. Soltis, R.P. Guralnick. 2019. Biodiversity synthesis across the green branches of the tree of life. *Nature Plants* 5: 11-13.

Potential Distributional Shifts Under Climate Change Models



Anson Wang



Ailanthus altissima
Casuarina equisetifolia
Centaurea stoebe* ssp. *micranthos
Dioscorea bulbifera
Lantana camara
Schinus terebinthifolia

Wang, A., A. Melton et al. 2022. *Plant Diversity*.

Time to Exercise!

What happens when there is a large dataset?

Should we manually create 40,000

Species Distribution Models?

We use BiotaPhy tools to automate

the creation of SDMs!!

Before we start ...

Download for the first time OR update the tutorials repository containing test data and configurations.

Initial download:

```
git clone https://github.com/biotaphy/tutorials
```

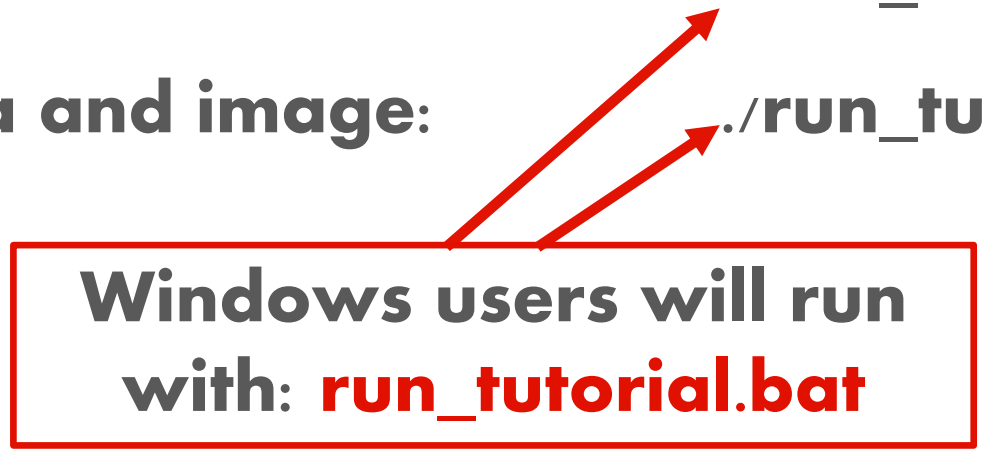
Update tutorial:

```
cd tutorials
git pull
```

Before we start …

Let's rebuild our Docker images to incorporate any updates. Move to the directory containing the tutorials repository that you downloaded or updated.

- ✓ Remove old docker elements: ./run_tutorial.sh cleanup_all
- ✓ Rebuild data and image: ./run_tutorial.sh build_all



Time to Exercise!

Let's put the automated framework developed by BiotaPhy to the test!

How to create Species Distribution Models (SDMs):

3 steps:

- ✓ **Data Preparation**
- ✓ **Run Tutorial**
- ✓ **Inspect Output**

Input: occurrence records

Input: environmental layers

Input: ecoregions file

Input: Script parameter file

Data Preparation: Occurrence Records and Ecoregions

Data Preparation

Input: occurrence records

The create_sdm tool accepts one or more occurrence CSV datasets, defined in two ways in the configuration file: either specified by a parent directory in the *points_dir* parameter, and/or a list of individual files in the *points_layer* parameter. The first line of each CSV file must contain fieldnames. Each of the occurrence datasets must use the same fieldnames for the species_key, x_key, and y_key, specified in the configuration file.

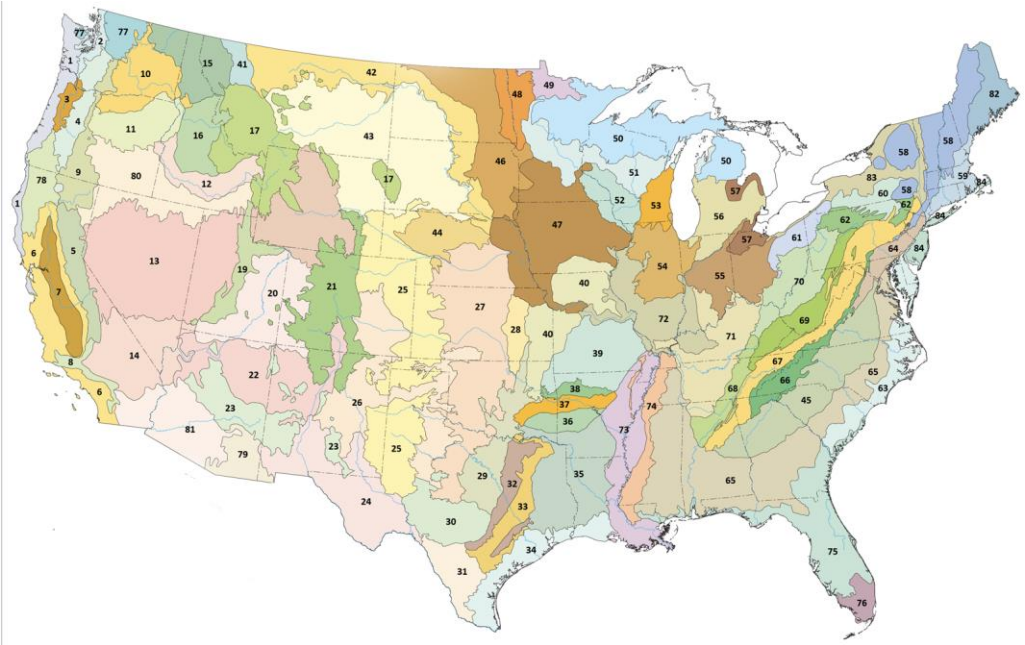
More information is in the **Occurrence Data** section of data_wrangle_occurrence.

Input: environmental layers

The environmental layers files are raster files of environmental data relevant to the species being modeled. Maxent, the only SDM algorithm offered in Biotaphy tools, requires environmental layers in ASCII format. Environmental layers are identified in the script parameter file by their parent directory.

Input: ecoregions file

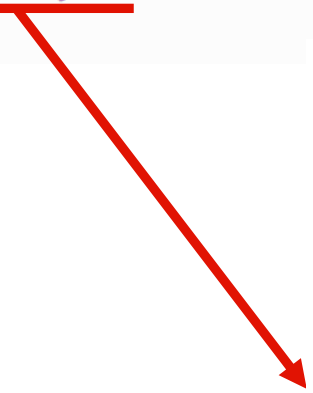
The ecoregions file is a raster file indicating broad ecoregions for the region being modeled. Ecoregions data can be used to compute a simple Rare Species Model for species without sufficient points for a Maxent model, by intersecting the convex hull of the points with the ecoregions. These data are available from coarse to fine scales, at global and regional extents. This file can be in ASCII or GeoTiff format.



Data Preparation: Script Parameter File

Input: Script parameter file

A JSON parameter file is required for this command. A test tutorial parameter file is:
create_sdm.json,



```
{
    "log_filename": "/volumes/output/create_sdm.log",
    "log_console": true,
    "report_filename": "/volumes/output/create_sdm.rpt",
    "min_points": 12,
    "species_key": "species_name",
    "x_key": "x",
    "y_key": "y",
    "points_dir": "/volumes/data/input/heuchera_rfolk",
    "env_dir":  "/volumes/env/na_2.5min",
    "ecoregions_filename":  "/volumes/env/ecoreg_na_2.5min.tif",
    "out_dir": "/volumes/output/heuchera_rfolk_sdm"
}
```

Data Preparation: Script Parameter File

- Required:
  - **out_dir**: Parent directory where the output species directories containing output data should be written. If the directory does not exist, it will be created
  - **env_dir**: Directory containing the environmental raster files for modeling and projecting species distributions.
  - **ecoregions_filename**: Raster file denoting ecologically and geographically defined regions to be used for modeling rare species or as a mask for the Maxent algorithm.

Required parameters!

```
{
    "log_filename": "/volumes/output/create_sdm.log",
    "log_console": true,
    "report_filename": "/volumes/output/create_sdm.rpt",
    "min_points": 12,
    "species_key": "species_name",
    "x_key": "x",
    "y_key": "y",
    "points_dir": "/volumes/data/input/heuchera_rfolk",
    "env_dir":  "/volumes/env/na_2.5min",
    "ecoregions_filename":  "/volumes/env/ecoreg_na_2.5min.tif",
    "out_dir": "/volumes/output/heuchera_rfolk_sdm"
}
```

Data Preparation: Script Parameter File

- **points_dir**: Parent directory containing occurrence data in CSV format. The tool will attempt to model all CSV files in this directory. Though this parameter is optional, one or both of *points_dir* and *points_layer* must be included.
- **points_layer**: List of filenames containing occurrence data in CSV format. Though this parameter is optional, one or both of *points_dir* and *points_layer* must be included.
- **species_key**: The field name of the column containing the taxon value in all occurrence data files. If this parameter is not specified, it will default to *species_name* (which is also the default value created in CSV data output from split_occurrence_data and wrangle_occurrences`

```
{
    "log_filename": "/volumes/output/create_sdm.log",
    "log_console": true,
    "report_filename": "/volumes/output/create_sdm.rpt",
    "min_points": 12,
    "species_key": "species_name",
    "x_key": "x",
    "y_key": "y",
    "points_dir": "/volumes/data/input/heuchera_rfolk",
    "env_dir":  "/volumes/env/na_2.5min",
    "ecoregions_filename":  "/volumes/env/ecoreg_na_2.5min.tif",
    "out_dir": "/volumes/output/heuchera_rfolk_sdm"
}
```

Optional parameters used

Data Preparation: Script Parameter File

- **maxent_params**: Extra options and parameters to be sent to Maxent. A full list of Maxent parameters, long with the value type, and sometimes a valid range of values, is available in the Maxent Github repository.

83 lines (83 sloc) | 13.1 KB

Raw | Blame

Search this file...

	Parameter	Abbreviations	Display text	Values
2	responseCurves	P	Create response curves	boolean
3	pictures		Make pictures of predictions	boolean
4	jackknife	J	Do jackknife to measure variable importance	boolean
5	outputFormat		Output format	string/Cloglog/Logistic/Cumulative/Raw
6	outputFileType		Output file type	string/asc/mxe/grd/bil
7	outputDirectory	o	Output directory	directory
8	projectionLayers	j	Projection layers directory/file	filedirectory
9	samplesFile	s	Samples	file
10	environmentalLayers	e	Environmental layers	filedirectory

Input: Script parameter file

- **min_points**: Minimum number of points in an occurrence dataset for Maxent to be used for modeling to. If the data contains less than the minimum, the *Rare Species Modeling* algorithm will be used.
- **log_filename**: Output filename to write logging data
- **log_console**: 'true' to write log to console
- **report_filename**: output filename with data modifications made by wranglers

Optional parameters used

```
{
    "log_filename": "/volumes/output/create_sdm.log",
    "log_console": true,
    "report_filename": "/volumes/output/create_sdm.rpt",
    "min_points": 12,
    "species_key": "species_name",
    "x_key": "x",
    "y_key": "y",
    "points_dir": "/volumes/data/input/heuchera_rfolk",
    "env_dir":  "/volumes/env/na_2.5min",
    "ecoregions_filename":  "/volumes/env/ecoreg_na_2.5min.tif",
    "out_dir": "/volumes/output/heuchera_rfolk_sdm"
}
```

Let's run this tutorial!

Run tutorial

with a list and a directory containing occurrence data files.

Initiate the create_sdm process with the following:

Goal: Create Species Distribution Models for multiple taxa at the same time

```
./run_tutorial.sh  create_sdm  data/config/create_sdm.json
```

```
./run_tutorial.bat  create_sdm  data/config/create_sdm.json
```

Remember, you will RUN this code in the terminal (Linux/OSX) or in a Command Prompt (Windows)!

Let's look at the output!

Output

Most outputs are configured in the script parameter file, and may include:

1. If "report_filename" is specified in the script parameter file, a summary of point manipulations by each wrangler will be written to this file.
2. If "log_filename" is specified in the script parameter file, that will be created.
3. If "log_console" is specified in the script parameter file, logs will be written to the command prompt during execution.
4. A directory named in the out_dir parameter, containing a subdirectory for each input occurrence data file. Each subdirectory will be named by the value in the grouping field and contain a predicted distribution raster in ASCII format. Occurrence data that were modeled with Maxent will also contain Maxent outputs. A subset of species outputs from the above command are in the directory heuchera_sdm.

Let's look at the output!

main / tutorials / data / input / heuchera_rfolk_sdm /

Go to file | Add file | ...

zzeppozz updated output — 4c67cb7 2 hours ago — History

..		
bensoniella_oregona	updated output	2 hours ago
conimitella_williamsii	updated output	2 hours ago
elmera_racemosa	updated output	2 hours ago
heuchera_abramsii	updated output	2 hours ago
heuchera_acutifolia	updated output	2 hours ago
heuchera_alba	updated output	2 hours ago
heuchera_americana	updated output	2 hours ago
heuchera_bracteata	updated output	2 hours ago
heuchera_brevistaminea	updated output	2 hours ago
	updated output	2 hours ago

/github.com/biotaphy/tutorials" in a new tab

Session Summary, Q&A and Discussion

- Species Distribution Models
- Fundamental vs. Realized Niche
- MaxEnt-predict potential geographic distribution (niche)
- Applications
  - Current, past, and future distributions
  - Phylogenetic Diversity (PD)
  - Polyploid nice vs diploids parents
  - Invasive
- Data Preparation
  - Input occurrence records, species list, ecoregions
  - SDMs for multiple taxa at the same time

Any questions??

Please use the chat to write your question!